

Visualizing Email Archives – *draft*

Judith Donath
MIT Media Lab
judith@media.mit.edu

Abstract

The email archives that people accumulate are a dense, complex, and highly personal record of their past interactions. As email becomes increasingly ubiquitous, these include not only their work interactions, but also their relationships with family members, friends, doctors, teachers, etc. However, current mail clients do little to support these archives beyond providing a basic filing and searching system: the interface to this very social environment renders it cold and impersonal. The problem we are addressing in this paper is how to transform this interface into a sociable space; a place where one feels surrounded by friends, conversations, and memories.

In the first section of the paper we establish the need for visualizing and interacting with email archives. We start by looking at email archive use today – how they are used, the problems that exist. In the second section we then look at what information exists within an email archive, and what part of it is amenable to computational analysis. We also look at the information that is contributed by the user and discuss alternatives to filing that would allow the user to organize email in a more intuitive and informative way. In the third section we describe and critique a series of existing prototypes, discussing the social patterns that are embedded in email, such as the rhythm of interactions, the flow of relationships and the social networks that connect individuals, and examine the significance of making these perceivable. In the fourth section we explore the metaphor of email as habitat: if email is a virtual home, how can it help its inhabitant thrive? Here, we look at how personal beliefs and histories are reified by objects in the real world and at how this can be accomplished with email visualizations in the virtual world; we examine the role of display in establishing social role and in creating personal narratives; we discuss how email visualizations can function as both a private record and public display.

1 Why visualize email archives?

Email has become a habitat[8]: it is one's virtual home away from home. People check email all day, from work, from home, and on the road. It is used to set up appointments, discuss politics, and begin romances. Although it can be used to send to-do items to oneself, etc., email is primarily social: it is a communication medium for connecting people.

A great deal of social information exists within one's personal email archives, though little of it is perceivable through the interfaces of today's email clients. Hidden among its cc's and bcc's is a depiction of one's social network, big changes in one's contacts demarcate moves to new jobs and new cities, and varying linguistic features and exchange rhythms create a portrait of one's acquaintances and their relationships.

This paper is about designing interfaces that reveal the social information contained within email archives. More deeply, it is about transforming these archives into a personal habitat, about using them to create a unique and meaningful virtual home.

Email is one of the most ubiquitous computer applications, and is arguably the most basic and fundamental form of computer mediated communication. It was invented in 1972 – and within a year three-quarters of all traffic on the ARPANET was email. Over the next few years the mail header protocols were established, and the basic functions found in today's mail clients were developed. It is useful to note that features that today seem fundamental and essential had to be invented – and were not obvious from the start. RD, a mail client written in 1973, was the first to separate messages for reading and deleting - before that, one had one long inbox file. In 1975, MSG, which has been called the first modern email program, was created by John

Vittal. It could automatically address replies, forward mail, and file messages into folders. MSG was designed to emulate physical mail; it included features such as cc and bcc, used today by millions who may have never seen an actual piece of carbon paper [16].

Over a quarter century later, today's email programs are not fundamentally different from Vittal's pioneering design. Yet much has changed around them. Email is no longer read at teletype terminals, but on personal computers capable of rendering complex graphics in realtime and of storing gigabytes of data. Computer and network access is no longer limited to a tiny group of American computer scientists but is a popular medium for millions of people worldwide. In 2001, an estimated 9.6 billion emails were sent each day [15].

One interpretation of this data is that email is perfect. It's been extraordinarily successful – why change it? Email is easy to write, easy to send. The basic ASCII format has proved quite versatile, usable for formal declarations and hasty notes, purchase orders and love letters. Its simplicity is its strength.

Studies of how people use email suggest that there is, however, room for improvement. In particular, problems with managing one's growing archive of emails are consistently cited as a concern. People have difficulty coming up with good categories for folders, they find it time consuming to file messages and once filed, often cannot find them easily when needed.

Beyond the practical problems of email management, however, are some deeper issues about what the email archives can become. In the words of Hallnäs and Redström: "When computer systems change from being tools for specific use to everyday things present in our lives, we have to change focus from design for *efficient* use to design for *meaningful* presence". Email has certainly become an everyday presence in our lives, but our tools are still only focussed on efficient use. The goal of this paper is to explore future interfaces for email archives that use this data to create a meaningful environment.

1.1 Revealing ties and context

In the physical world, we are accustomed to an environment replete with detail. We read a great deal of information in our companions' facial expressions, clothing choices, tone of voice. Our recollection of social patterns, such as who knows whom, is aided by our memory of places and events. We live amidst an interlocking series of temporal patterns such as the changing of the seasons, the cycle of weekly management meetings, our daily routine; our sense of social patterns may be measured against these rhythms.

Email, by contrast, is quite spare. There is little visual detail to distinguish messages or contacts, and many socially relevant details, such as co-location on a cc list, are easily overlooked. Email flattens space and time. An email from a colleague across the hall appears no different than one from across the world. It is asynchronous – I can write to you whenever I wish to, and you will read and respond as you can and wish. Online, it can be hard to keep track of who has been privy to the same conversation, to remember when you last spoke with someone, or to have a clear sense of how frequently you communicate with particular person. Today's email clients may provide us with an automatically generated list of contacts, but such a rolodex is missing the rich contextual information that helps us sort through and comprehend our social environment. There are significant temporal rhythms in email [37], but they can be difficult to perceive.

Visualizations can provide some of the missing context by revealing the data and patterns that are hidden within the email archive. They can show when you met someone, what are the connections among your acquaintances, major shifts in your contacts. At the organizational level, they can show who works together, who is the link between groups of people, how do new employees become integrated into the company.

Such visualizations are practical and useful. They help people manage their expanding network of correspondents and growing collection of messages. Beyond the immediately pragmatic, they allow people to observe and reflect upon their online social world.

Email has radically changed the scale of our social world. We are likely to be in personal (though virtual) contact with hundreds or even thousands of people, a far higher number than was previously common, or indeed possible. The numbers alone indicate a need for help in keeping track of these contacts and interactions; the need is exacerbated by visually sparse and abstract nature of online interaction.

Robin Dunbar has proposed that language evolved because grooming was an inefficient way to maintain social ties. Now it appears that email may make face to face communication seem quite inefficient. A recent Pew Internet study shows the most popular online social activity of college students was to forward email to friends. Such notes are a form of social grooming: their pre-made content, like that of postcards, is not the point of the message, which is simply to stay in touch; the pre-made content lets one do so quickly and easily. The Pew study showed that today's college students were staying in touch with more of their hometown friends than did their predecessors. Before email, the effort to stay in touch was greater and when faced with the big new social circles of college, the time and effort to keep up a larger circle of friends was too great. But today, with the efficient mechanism of email, maintaining social ties is easier and less time-consuming. This widening of the social circle, in conjunction with its maintenance via the sparse medium of email, underscores the need for tools to help keep track of and understand one's social environment.

Dunbar also predicted that electronic mail would not enlarge people's social networks. The human mind, he said, has a limited ability to handle information about other people. Yet tools have always extended humankind's reach beyond our native abilities. Here, we will be looking at the tools that we need to thrive in today's highly networked and connected world.

1.2 How do people manage email archives today?

There have been numerous studies of how people use email ([7][8][25][27][37][41][42][43]). They paint a picture of a communications medium of increasing ubiquity, one that is used not only for communications, but for task management, personal archiving, and contact maintenance. Broadly speaking, email seems to function quite well as a communication medium, but current tools poorly support the management of email archives

MacKay [25], Whittaker and Sidner [43] and Ducheneaut and Bellotti [8] observed that there is a wide range of approaches to handling large amounts of email. Some people scrupulously delete unnecessary mail, while others keep almost everything they've received. Some file email into folders, typically organized by topics such as sender, project or hobby, in numbers ranging up into the hundreds [8], while others do not file at all, keeping a single large inbox, sometimes archiving this inbox by date when it gets too large [25]. People said they kept email messages for a variety of reasons: as reminders of tasks to be done, as records of conversations and decision-making processes, and a history and list of contacts.

It is important to note that most of these studies of email use have been conducted in work settings: the email accounts that were analyzed were work accounts and the subjects were interviewed in their offices (in one study [43], the subjects' company was also the developer of the email client). It is likely that this setting caused the respondents to emphasize the pragmatic reasons for their behavior, to say that they are saving vast quantities of correspondence for its potential information-retrieval uses rather than for a perhaps inchoate desire to collect and save their history, no matter how mundane.

Whether practical or sentimental, the function of email archives to remind was frequently mentioned, not only the immediate reminding of a task to be done, but of a relationship to be maintained, a thought to be continued. In the course of their daily work, people often reviewed portions of their inbox (which might have hundreds or thousands of emails) as they scanned for a particular piece of mail, or simply looked over the list of contacts. Although all email clients provided search tools, people preferred to use sort functions to find particular pieces of email [8], a strategy that, deliberately or not, helps email work as a memory tool by refreshing the user's familiarity with the archive.

Today's email clients are all designed to manage email archives by filing messages into folders. Yet filing is cognitively difficult and almost all users expressed difficulty with this approach. People had difficulty with generating appropriate folder labels and with remembering these labels later for filing and retrieval [43]. If one has a folder for sender X and for project Y, where should a message from X about Y be filed? People had redundant folders, folders that were too big or too small, and folders that had long since ceased to be used. People worried that filed emails would get lost, and that filed junk would be kept forever. Once an email is filed, it no longer serves the purpose of reminding. We address the problems of categorizing and sorting email with a proposal to eliminate folders for email archives, and instead marking emails with user defined attributes (discussed in detail in section 2.3). While the focus of this paper is on

visualizations, the process of filing or marking messages is the primary way that users annotate their email, providing data that is quite useful for creating meaningful visualizations.

A key question that these studies raise is why do people save so much email? Although some people scrupulously delete unneeded mail¹, many save quite a lot of mail, including some who archive almost everything². People's responses when asked why they saved so much often indicated a fear of deleting something that someday they would want. When pressed, the responses were pragmatic – email might contain some facts or commitments that would need to be referenced later. Yet that does not seem to address why so much seemingly mundane correspondence is saved, e.g. messages planning meetings or lunches, casual exchanges, etc. Our hypothesis is that people want to save the history of their experiences, achievements, and relationships. They value the accumulation of exchanges, in which the individual messages may not be very significant, but when considered as a whole, creates a meaningful pattern.

An important caveat is that these studies have been conducted on email as it exists today. So while some people are meticulous about discarding old email, that behavior is occurring in a context in which there is little reason to keep lots of old email, and many reasons to throw it away (for instance, to keep the mailboxes from becoming too unwieldy). If this mail contributed to the detail and accuracy of a personal email visualization, they may well choose to save them instead.

2 What information is in email?

Any visualization is constrained and determined by the data it is given. With email, the data can come from three sources. The header information is highly standardized and computationally accessible. The message content is free-form text: linguistic analysis could derive some information from it, but it is very context-dependent. Finally, there is data supplied by the user. This is likely to be easily accessible by a computer, the question here being the user's motivation to annotate. Here, we look at these three types of data in greater depth.

2.1 Header

The header indicates who the mail was sent to, whether just the user or a long list of cc'd recipients. It tells who it was sent from, the time and day it was sent, and basic threading information.

This data provides sufficient information for a number of interesting visualizations. The sender and receiver data (to, cc, bcc, and from) provide the foundation for drawing a network of the connections among all the contacts. The temporal data allows us to depict historical trends. Further interpretation of these bits of information can yield more nuanced views – a person whose mail I receive only because I am on a mailing list and never through personal contact has a different place in my personal social universe than someone with whom I am in direct contact. We can detect patterns of reciprocity, and analyse the rhythms of conversations.

The header information is easily accessible and computationally tractable. For many people, having a program that creates visualizations from analyses of email headers is acceptable whereas one that analyzes the contents is not. The prototype visualizations we discuss in section 3 are primarily based on data from the header, along with some user supplied information.

¹ Some users are impelled to delete mail because of space limitations; presumably, they would save more if storage were not an issue.

² It is becoming an increasingly common practice for companies, especially in fields such as finance, to require that employees save all their email, as is legally required for them with traditional written correspondence. The subjects interviewed in the studies reviewed here were not working under such a requirement.

2.2 Content

There is an enormous amount of potentially interesting information in the body of email messages. There is semantic information: what is the topic being discussed? There is affective information: what is the tone of the message? And there is social information: what is the relationship between the writer and recipient and what are their respective roles in the larger community?

Semantic analysis allows us to group related messages and people, display the ebb and flow of particular themes, and create a context for the display of other information. There is considerable research in the field of information retrieval on algorithms for determining the topic of a text and, increasingly, on visualizing the relationships among a set of documents (e.g. [46], [33]). Semantic analysis can range from simple keyword detection to complex analysis incorporating a thesaurus and analyzing the relationship among the words.

Affective analysis provides insight into the personality of the writer and the relationship between writer and recipient. Work has been done to automatically detect anger and insults in newsgroup messages [35], but detecting other emotions is more difficult. Anger in email and newsgroups is often expressed using a limited vocabulary of hostile words, phrases and sentence structures, while supportiveness, confusion, sadness, etc. are more subtly indicated, and even human readers may disagree about the affective intent of a message.

Social analysis attempts to infer information about the roles of the sender and recipient, and their relationship. Panteli [27] has examined the power relationships that appear in email, looking at such features as the variations in formality in messages sent up or down the organizational hierarchy, while Duchenaus[7] has examined the way email can be used to reinforce (or subvert) the structure of an organization. Both of these studies used human readers to analyze the content of the messages, extracting information that is far beyond the ability of today's computers: Duchenaus, for instance, coded messages for sarcasm, anxiety, formality, fear, among other features. However, he also coded for a number of features that could be automatically extracted, such as citations, non-grammatical writing, the use of the subject line, etc.

Warren Sack's "Conversation Map"[33] is an interesting example of a visualization that combines both semantic and social data to create a map of a large scale conversation (Usenet newsgroups). He uses patterns of response, especially mutual citations, to create a social network of the participants. [describe semantic network] to create a map of word usage particular to that conversation. *Conversation Map* allows the viewer to explore the social and semantic space of the conversation. It is worth noting that here the content of messages is examined within the context of the entire corpus of the conversation – the point here is not to find the meaning or importance of an individual message, but to map the social groupings and textual meanings that shape the overall discussion.

In comparison with the header data, information from message content is harder to extract computationally, and is likely to be far more subjective. This material is also more private than the header, which is important when we discuss public vs private displays of email visualizations. Yet it is also a very rich source of information about the meaningfulness of the email archive, whether it be cues to the subject matter of conversation or indications of the type of relationship between two people.

2.3 User annotation

Some of the most important information about a piece of email exists nowhere in the message itself – it is the meaning that the communication has for the recipient (or sender). Seemingly innocuous notes may be souvenirs of momentous occasions: the significance of a message that says "Can you make it to lunch at 12:30 today" depends on the events at that lunch. The meaning of a message may depend on the relationship of the sender to the recipient, a relationship that may well have subtle yet important distinctions. How we think about our own email – about the significance of the message and our relationship with the sender – is an important source of data for visualization. The key question here is how to obtain this information from the user.

Users currently do some annotation, in the form of their choice of folders in which to file the message, and, with some programs, an additional binary flag or choice of one of a limited set of labels. Unfortunately,

this categorization structure is both limited and rigid. As an annotation, folder placement provides minimal data. Worse, users find the process of filing to be cognitively difficult, and many avoid doing so altogether.

Our recommendation is that the email client be redesigned so that instead of filing messages into folders, users would mark messages as having any of a number of user-defined attributes. Not only would this provide more detailed and extensive annotation data, we argue that it would be cognitively easier than the current method.

Whittaker and Sidner[43] noted that people found filing email to be cognitively difficult. They had trouble figuring out what to label their folders, they had difficulty deciding which folder a message should go in, and once filed, they were concerned that the message would become “lost” – that they would not remember into which of several possible appropriate folders they had placed it. Creating good folder categories is indeed difficult: one needs them to be broad enough that a sizeable selection of messages can be placed in them, but not so broad that they become unmanageable. Most problematically, they should ideally not overlap; in a perfect (and in most cases, impossible to design) system, each message would belong in one and only one folder. Yet people use diverse and incompatible categories – they may have a folder for each project they work on, and also a folder for mail from particular co-workers. Where does a message from one such colleague about one such project belong?

If we think of these categories as attributes with which to mark messages, rather than as folders in which to place them, then we have a much more flexible system. A message can be marked as being from a certain colleague and as pertaining to a certain project; it can be additionally marked as containing budgetary information, as being confidential, as having a funny joke included in it, etc. Each message can have as many attributes as suit it, and that the user chooses to add.

Leaving aside, for the moment, the goal of visualizing email archives, we note that the attribute marking approach could be integrated with an interface similar to today’s clients. Attributes would be listed at the side of the screen (as folder names are now) and choosing an attribute would bring up a list of all messages with that attribute (just as clicking on a folder name today does). The display could be expanded or narrowed using Boolean combinations of terms. Determining one’s sets of attributes would be easier than determining folders, for they need not be primary or non-overlapping categories. User’s would no longer need to decide into which one of several appropriate folders to place a message – any number of attributes could mark it. And, there would no longer be the concern that a message would be “lost” by being misfiled, for it could be found by any of the appropriate features.

An attribute system is well suited for automatic processing. Many email programs today have the ability to automatically file messages, though most people don’t use this feature, or use it only to sort out listserv email, etc. stating that they fear that if the email is filed before they read it, they will never see it.[8]. Automatic feature marking, however, would not have this drawback. Attributes such as the sender could be easily marked automatically; indeed, any of the data discussed above that is computationally extractable could be used as an automatically marked feature. As Duchenaut and Bellotti point out, such filtering should be applicable at any time, not only upon new messages (as is the case with several contemporary clients).

This is just a brief outline of an attribute based email classification system. Implementing one would require working out a number of design questions. One is how to use temporal data to limit the display: the full mail archive will quickly become unwieldy, even when viewing only matches to selected attributes, so a temporal bounding function would be quite useful. Another is how to design ontology for structuring the attributes: we have thus far described attributes as basic keywords, but more powerful (though complex) structures can be useful. One may want to have categories for different types of senders (friends, co-workers, etc.), and to be able to label someone as a member of multiple categories (e.g., a friend with whom one works).

Thus far, we have been talking about deliberate message classification (whether through today’s filing process or a attribute marking process) as the source of user annotation. Another source is observing how people interact with a message. Today’s email clients note whether one has read, answered or forwarded a message. Potentially, *all* of one’s actions can be considered data sources. Hill, Hollan et al [21][22] proposed recording all of one’s interactions with a document and then visualizing, as an integral part of the

document, how it was read, edited, etc. With *EmailWear*, one of their experiments in this area, they recorded how much time people spend reading each message and how they navigated through their mail. The idea here is to think of all actions contributing to an understanding of the meaning and significance of a message.

2.4 The data in email

There is an enormous amount of data to be found within one's email archive, both in the messages themselves and in the way they are handled by the user. Some is very easy to extract, some takes considerable computational sophistication, and some requires rethinking the user interface to encourage the user to provide the desired annotation. Some of the data may be thought of as relatively innocuous, whereas other information may be very private. Some data may be factual and objective, such as the date on which a message arrived, while some may be quite subjective, such as the interpretation of the tone of a message. As we create the visualizations, it is important to keep these qualitative distinctions between different types of data in mind. A public display should not be based on data the user wishes to keep private; subjective data should not be rendered in a way that makes it look like objective fact.

In the next section, we discuss a series of email visualization prototypes. These initial works are primarily based on information found in the header. We conclude the section by discussing future directions, including the incorporation of message content based social and semantic data.

3 Email visualizations

There are many possible ways to visualize email. Time can be the organizing principle, with the visualization revealing the polyrhythmic cadences of project deadlines, relationship evolutions, weekly work patterns, etc. Or social relationships can be the primary focus, with the visualization exploring how the correspondence patterns weave social networks. Or topic can be the key, with the visualization providing an interactive map of the projects, ideas, people and places one has discussed.

Visualizing email is a recent concept and to date there are only a few prototype systems. In this section, we discuss four of them, two that use time as the organizing principle and two that explore social networks. *Post-History* and *Social Network Fragments* were developed in our research group; the other projects, elsewhere.

3.1 Time based visualizations

There are many way in which men are made aware, or rather make themselves aware, of the passage of time – by marking the changing of the seasons, the alterations of the moon, or the progress of plant life; by the measured cycling of rites, or agricultural work, or household activities; by the preparation and scheduling of projected acts and the memory and assessment of accomplished ones... But surely among the most important is by the recognition in oneself and in one's fellowmen of process of biological aging, the appearance, maturation, decay, and disappearance of concrete individuals.

- Clifford Geertz "Person, Time and Conduct in Bali"

Although email is an asynchronous medium, it is not arrhythmic or atemporal. There is a tempo to email conversations: there are fast exchanges of one line notes, a rhythm approximating written speech, and long slow correspondences more reminiscent of the unhurried rhythms of letter writing in the days before cars and planes.

Tyler and Tang found that people notice the rhythms of others' responses and are aware of the impression they form with their own [36]. They felt obliged to respond quickly to those who were prompt in answering them and were inclined to postpone answering the emails of those who answered more slowly. People may wish to seem as if they are working hard and thus be diligent about answering quickly; this can be especially important if they are working remotely and want to avoid the appearance that they are not working at all. Sometimes, people may wish to convey the impression that they are not easily accessible,

and may deliberately delay responding so as to appear too busy to respond immediately. The temporal mores for responsiveness can vary from organization to organization, person to person, conversation to conversation.

Email users live amidst numerous unfinished conversations. At any given time, they may have tens (or, for some, hundreds) of messages awaiting their response, and many others for which they are awaiting an answer. A tool that highlights mail that has gone unanswered and displays it in the temporal context of previous exchanges, would be useful as a reminder to respond (or to prompt for a reply) and maintain the desirable response time.

Current email clients allow users to sort their mail (within a single folder) by date; otherwise, they provide little indication of the temporal nature of email. We will look first at two experimental email visualizations that depict temporal rhythms, and then discuss further work to be done in this area.

3.1.1 TimeStore

TimeStore (1994) [45], was a redesign of the email inbox and filing system which automatically arranged email in a two dimensional grid, with one axis being people and the other, time. It helped the users see patterns of correspondence and to note where there was an unintended lull. [image] Each message was represented as a dot in the time/person grid, thus individual conversational rhythms were easy to discern, but the subject of each email was obscured.

Preliminary tests of this project yielded some interesting observations, most notably that people liked the time-based display, especially the ability to see trends in their correspondence with friends and colleagues. However, they did not find it usable as a standalone mail interface, for it requires knowing the arrival time of an email in order to find it, and users often needed to open numerous messages to find the one they were looking for. The interface made people think of the system as more of a database than is a regular mail client, and they therefore wanted it to have database-like abilities, such as the ability to sort the view by how much time one spends corresponding with someone.

TimeStore demonstrates that a simple visualization of email receiving patterns can be quite useful. The difficulty with finding a particular piece of email could be addressed by including subject lines in the visualization (perhaps in conjunction with a zooming or lens function to avoid over-cluttering the screen). The users' desire to have more database functions confirms that the exploration of various patterns in email, many of which are completely inaccessible in today's clients, is a useful direction to explore.

3.1.2 Post-history

Post-History also visualizes the rhythms in one's email conversations [39] Here the emphasis is on depicting overall patterns of email activity and the changing prominence of particular interactions.

The *Post-History* display has two panels, one showing a calendar and the other, one's contacts. [image] the calendar are shown as colored squares, whose size and brightness varies to show quantity of email sent and received (size) and how much of it was directed immediately to the user vs. received via a distribution list (brightness). The contact list shows all of one's contacts (within a large, bounded time period). The most active ones are shown at the top; the overall shape of the mountain of contacts indicates one's range of conversational patterns. Users can see the rhythm of their conversation with a particular individual by clicking on that person's name, which causes yellow squares to appear on the calendar for each message received from that user, and white squares for each message to that user.

Post-History animates the changing patterns of frequent correspondents: the display can be played as a movie through time, and one can watch names move up and down as conversations wax and wane.

Participants in a small-scale user study were very interested by the patterns *Post-History* revealed in their email. They found it interesting to see who they maintained steady contact with, who had been a frequent correspondent for a brief but intense time, as well as simply the quantity of people with whom they are in contact. Some mentioned that they had not been aware how transient many of their interactions were. Others noted the overall temporal rhythms their email revealed: the Monday morning flurry, the relative solitude of vacations, the buzz of discussions during intense project development times.

Post-History is a work in progress and within the existing framework a number of additional patterns could be featured. For example, in the animation of the contacts, using a different color for new contacts and deepening the color over time would provide an interesting depiction of the duration of one's online interactions and would show transitional periods (such as a new job) that involve developing many new relationships.

3.1.3 Time based email visualizations

TimeStore and *Post-History* are but two of the ways that the temporal aspects of email can be visualized. Between these two interfaces, the three basic formats for conveying temporal information are represented: the calendar, the timeline, and the animation. Both of these interfaces used as data just the sender and the date; supplementing this basic structure with other information could greatly enrich the visualization. Additional information could be as simple as the subject line (both to assist in finding messages and to see the changing (and perennial topics) , it could be user supplied annotations, or various analyses of the content itself. Both of these focus on dyadic email exchanges; by analyzing the co-recipient patterns (see below) the evolution of one's social network could be depicted.

Users of *Post-History* said that they would like the ability to annotate some of their history, to explicate the events that caused anomalous patterns or big changes in their contacts. Similarly, the designers of *TimeStore* decided to supplement later versions with calendar and other event markers. These annotations – are they for the user, or for observers? One theory is that the temporal visualization is indeed analogous to a scrapbook, that is, a collection of items from the past that is meaningful to the owner, and also shareable with others. We will discuss this notion in greater depth in section 4

The time-based email visualizations use email as personal history, like an extended diary/scrapbook that contains elements from all aspects of one's life. Today, most adults have only a couple of years of email archives, if that, and these may be quite incomplete. But today's schoolchildren are already accustomed to communicating electronically with their family and friends, perhaps building the start of an archive that will help them trace the trajectory of their life as they go through high school and on into the world.

A key question with the temporal email visualizations is how to manage long spans of time. One's list of contacts grows throughout life, for one never "unknows" someone, and our social milieu grows more complex over time. Yet the number of active relationships may stay relatively stable, for others ebb over time (see Burt [4] for a discussion of why certain relationships decay faster than others). Colleagues from a past job with whom you have not stayed in contact fade into the past (indeed a useful function of an extensive email archive would be to provide you with a reminder of who they are when one unexpectedly contacts you 10 years later). An important function of a temporal visualization is to allow us to see this growth and evolution. In the next section, we discuss the visualization of social networks – clearly, there is a useful combination to be made showing the evolution of these connections.

3.2 Social network visualizations

A second approach to visualizing email is to depict the social networks that emerge from an analysis of the contact information. The notion here is to draw a picture of how a group of people are connected in order to understand who communicates with whom, who is the link between otherwise disparate people, and who is a central figure in a network of connections.

Social networks may be depicted in the first person, from the point of view of a single member (aka ego-centric networks), or in the third person, showing the entire network (known as whole network analysis). Analyses of someone's email archive produces an ego-centric network , showing that person's contacts and the connections among them of which he is aware³; there may well be connections that exist between two of his acquaintances that are unknown to the subject. *Personal Map* and *Social Network Fragments*, discussed below, are examples of email based ego-centric network visualizations. To create a whole

³ By "aware" we mean that there is some indication of awareness in the data, e.g. co-presence on a cc list, etc. The user may well not be consciously aware of these connections; making them more apparent is in part the value of these visualizations.

network visualization, a database with information about everyone's contacts (at least within a particular domain) is needed. *Visual Who* is an example of an email based whole network visualization; it uses an organization's mailing lists to depict the members affiliations from a third person perspective. Note that unlike the ego-centric models, there is no inherent center to this network, and thus a more highly interactive interface is appropriate (in this case users can choose particular topics to act as anchors for a spring based visualization). There are also hybrid visualizations. *BuddyZoo*, a visualization of instant messenger buddy lists, takes individual contact lists, which themselves have no inter-contact connection data, and links them to build an expanded and interconnected network.

Within the scope of this paper, the ego-centric network is the most relevant, as it can be derived from the user's email alone. It is important to note that these views are both subjective and incomplete – the actual social structure may be quite different from what is represented in a single user's email records. We will look first at two experimental email visualizations that depict personal social networks, and then discuss what can be learned from these and the further work to be done here.

3.2.1 Personal Map

Personal Map depicts the user's email-derived social circle, placing the user in the center, surrounded by his or her contacts, arranged in groups based on their co-appearance in emails and radially arranged so that the most "important" contacts are nearest to the center. [image] A small user study found that most people enjoyed using it and found that the groupings to be fairly accurate, though the ones that were based on small similarities seemed erroneous to the users. When asked how they would use it, the top reason was "for fun", followed by as a tool for sending group emails.

Personal Map was partially integrated with Microsoft Outlook, and was used to generate names for an add-on related people list. Users liked the idea of the list, which is meant to be used to help compose mail to groups of people, however, they did not like that it changed dynamically, for they expected names to appear in the same place and the dynamics disturbed their spatial memory. This is different from reactions to the dynamics of the *Personal Map* groupings, which users found "fun" and interesting to explore to discover who is where in their social network. Some of the difference may be due to the visual design: the colored blobby shapes of the *Personal Map* group space seems inherently fluid, while the black and white list of the *Outlook* interface suggests a static layout. The different uses are also likely to promote different expectations: when using a visualization to explore, dynamic changes are perceived as yielding interesting patterns, but when using a tool to compose email, these same changes seem annoying.

3.2.2 Social Network Fragments

Social Network Fragments is a depiction of the one's social network as it is woven by message addresses. It displays all the names in the network, arranged according to the closeness of their ties to each other. (see figure) The data is analyzed at timed intervals, so one can watch the evolution of the network – when people appear, when they become linked to others, when they fade into the background.

The program analyzes a person's entire email archive, and derives tie strength among the contacts from their co-occurrence in the address. In its initial analysis, it makes an interesting distinction among different types of ties. For creating the visualization, this distinction is reduced to different degrees of tie strength, but is worth describing as an interesting example of the nuances that can be made in network description.

- A sends mail to B: A *knows* B
- A receives mail from B: A is *aware of* B
- A and B both receive a message from C: A and B are *weakly aware* of each other (and are both *aware of* C)
- A receives a message via a listserv from B: A is *listserv aware* of B
- A blind carbon copies a note to B: A is *aware of* and *trusts* B

Weights are assigned to each of these types of ties. The email record is analyzed in two week intervals and visualized using a spring system algorithm: the individual nodes repel each other, while they are

simultaneously being pulled together by a lattice of different strength links. The system settles when it has minimized the distance between tightly bonded individuals while maximizing the distance between unrelated individuals.

SNF also colored the names of people based on classifications made by the user. This has the effect of overlaying the social network structure derived from the email data with the affiliation patterns perceived by the user. This overlay helps highlight particular groupings – some as very homogenous, others quite heterogeneous. *SNF*, in its current version, does not allow for interactive reassignment of the classes and colors, but this could certainly be implemented. Being able to see what different patterns emerge through the interactive overlaying of different affiliation groups (or, indeed, mapping any attribute to a color) is a very informative way of exploring the data [6].

3.2.3 Visualizing social networks

Both *Personal Map* and *Social Network Fragments* derive their depictions of the social network from contact names in one's email, grouping them according to their co-occurrence in address fields. They differ in their algorithms for creating the groupings and drawing the network, and certainly there are other algorithms and designs that could be used to visualize this data. The essential question, however, is not so much whether a particular analysis is optimal (though that is important too), but what does this data really tell us – and how can we use it?

Like any visualization, the accuracy of these social network depictions is limited by the data upon which they are based. How representative is the email addressing record – varies from person to person and from organization to organization. There are technical limitations: if the program does not have access to the membership of mailing lists, it will be inherently incomplete. At the Media Lab, for example, it is quite easy to create mailing lists, so many groupings may be addressed via such lists, rather than via cc'd addresses. Some lists have open rosters, but many others do not; it is reasonable to assume that almost any email visualization will be missing some group membership. Furthermore, lack of co-presence in my email archive is far from a reliable indicator of lack of connection. There are certainly people among one's acquaintances who are "linked" in some way who do not co-occur on any emails. The address-based ego-centric social network structure will generally be incomplete.

Indeed, what is interesting is how well it seems to work, even with these inaccuracies. User feedback for both *Personal Map* and *SNF* was positive, with the users seeing meaningful and recognizable patterns in the groupings. The people that showed up near each other generally seemed to be the ones one would expect to see together, and when they were not, often a forgotten or unexpected tie among them could be deduced. The visualization thus both affirmed the user's existing model of the social structure and extended it.

Both *Personal Map* and *SNF* used tie strength to indicate differences among the connections between people: the former mapped message frequency and the latter a combination of message frequency and type of addressing. Yet is tie strength the right mapping for this data? The stronger the tie, the closer the nodes (people) will be pulled, and the closer their relationship will appear. Are we losing valuable information by conflating trust, acquaintanceship, and awareness into the single dimension of tie strength? How well does frequency of contact map onto the importance of a relationship? Many leaders choose to keep themselves aloof; an email from one of them is a rare and valuable item. A visualization that looks only at frequency will equate the 2 or 3 exchanges one has had with the director of the company with the 2 or 3 exchanges one had with say, a distant colleague's assistant. This is particularly problematic with individual ego-centric networks. Studies that have found organizational structure to be replicated in email (e.g. [36]) were looking at data from a whole network. A central figure might be in contact with a wide range of people, but each one only occasionally; this centrality would be invisible to each of those contacts, who would see only their individual data.

We also need to be careful of the assumptions we draw based on these records. For instance, *SNF* assumes that if A bcc's B, then A is aware of and trusts B. This is generally true – if A sends an email to C and bcc's B, this scenario often means that the email was for some reason sensitive, and A trusts that B will not reveal this secret copy to C. However, bcc can also be used for just the opposite reason. If I am sending a notice to a large group of people – a product announcement or a party invitation – I may use bcc because I

do not want the people on the list to know who the others are: I may be concerned that someone might mistakenly (and annoyingly) reply to the entire group or worse, use the list to harvest email addresses. Here bcc is an indicator of lack of trust. As visualization designers we need to know when to simply depict the data and when to interpret it. If the data is difficult to perceive, simply depicting it can make it more accessible. Interpreting the data (as when we interpret bcc to mean “trust”) potentially can make the visualization more useful – the added semantic information can be used to highlight important data and to make the image more intuitively understandable. But it comes with the risk that the added data is biased or inaccurate; the tradeoff is often between legibility and objectivity.

What is the next step in visualizing social networks? One possible direction is to incorporate some of the differences among people and among their connections. Maps of social networks are impoverished representations of the rich, complex social world. They consist of nodes, which are people, and links, which are the connections between them. Neither are represented as having any special qualities, other than the differing strengths of the ties. Yet in the email data there is potentially a great deal of information that could be introduced here – from responsiveness rates to inferred hierarchies. To figure out what further data we wish to show, we need to have a goal for the visualization – what are we trying to achieve with it. In the next section, we propose the goal of making email into a richly meaningful virtual habitat – place for both reflection and connection.

4 Email as habitat

Email is often thought of as a tool. The makers of email programs design for corporate customers – not the individual users, but for purchasing decision makers, whose goal is to increase productivity and efficiency. Yet when we look at the way that email is actually used – when we see that millions of people use it for personal correspondence, when we see that even within the work context it has become “overloaded” – it is clearly far more than a tool. It is the place where computer users spend much of their day, it is the place where mementos are kept and relationships formed. In the words of Ducheneaut and Bellotti, email has become a habitat [8].

Whether we think of email as tool or habitat has implications for how we design visualizations – for how we imagine people using them, for how we evaluate their usefulness, for the questions we expect them to answer and the patterns we have them reveal.

When we think of email as a tool, we expect email visualizations to help us solve practical problems. We want to be able to see all our to-do items at a glance, we want to quickly notice where a conversation has faltered, we want it to tell us who to go to with a question we have. These certainly very useful things for an application to do – and the visualizations we have discussed can indeed help solve these problems. Yet these do not address the problem of how to design for a habitat.

A “habitat” is a place where something is commonly and naturally found. It can be friendly, a place where the organism thrives, or it can be unfriendly, making life harsh and unpleasant. For an animal, a good habitat keeps it safe and comfortable, able to find food and avoid predators. Humans too seek these necessities in their surroundings, but they also search for more.

In *The Meaning of Things*, Mihaly Csikszentmihalyi and Eugene Rochberg-Halton examine how people give meaning to their homes, observing what objects they find important and why. “One of the most important psychological purposes of the home is that those objects that have shaped one’s personality and which are needed to express concretely those aspects of the self that one values are kept within it... [it is] a shelter for those things that make life meaningful”[5]. Their findings provide a useful framework for thinking about email as a habitat, one that not only allows the user to function efficiently, but also to inhabit a meaningful virtual home.

Csikszentmihalyi and Rochberg-Halton interviewed many families to find out what objects were meaningful to them and why. They found that a meaningful object may evoke memories of a person, a place or an important event. It may represent a significant aspect of someone’s life, such as a skill they have acquired or an achievement. It may be valued for its intrinsic qualities: for its utilitarian function or for the enjoyment that its use or appearance bring.

These categories of meaning are also applicable to email. There are emails that evoke memories, whether because they recount an incident or are the corpus of notes exchanged with a particular person. The collection of emails about a project commemorates an achievement. There are emails that are valuable for the stories they contain – these are enjoyable in their use; there are those that are valuable for the information they contain – these are functional.

Csikszentmihalyi and Rochberg-Halton observed that the objects that people found to be meaningful varied considerably between men and women, and among children, adults and older people. When asked what were the most meaningful objects in their homes, women most frequently mentioned objects that signified personal relationships, men indicated objects that conveyed status and achievement, and children were most interested in things they could use, such as toys to play with and beds to sleep in. Studies of email usage show parallel behaviors. Children’s interest in the internet is primarily for entertainment (it is a toy to play with), transitioning to the more adult pattern of social engagement and information seeking around college age; and women use email as a social medium more than men do.

An email interface needs to consider these differences. For some users, the most important material might be the personal narrative of relationship trajectories; for others, it may be the achievements represented by projects completed or sales closed. The material one wants depicted, the style of the depiction, and the audience for it may vary not only from person to person but even for a single user from context to context: the types of things that are significant in one’s work context are quite different from the those in the personal sphere.

4.1 Cultivation

Thinking about email as a habitat rather than solely as a tool changes the way we think about interacting with it. A tool should be efficient; it should get the job done fast and unobtrusively. A habitat, however, can be a place for contemplation, for recalling events, thinking things over. Using a search function is using a tool; reading through the inbox is a form of inhabiting it. It is interesting to note that even with today’s mail clients, which have fairly powerful search features and are not very well designed for browsing, people prefer to look through folders to find what the message they are seeking rather than use the search mechanism [8][43].

Considered only from the standpoint of efficiency, it would seem better to use a search function rather than perusing the inbox to find a message – it is quicker, with fewer opportunities for distraction. Yet perusing the inbox has its own important functions. According to Whittaker and Sidner “people use external environmental sources or artifacts to mediate cognition... email folders function as an external archival memory store... people frequently access the email inbox which means that properly organized it can operate as a visual device for attention manipulation and reminding, as a way of extending working memory, and maintaining the context of ongoing activities in an interrupt-driven environment.” Other users reported perusing the inbox to remind themselves of who they knew; it is a social memory aid.

Thus, one important function of email visualizations is to provide a place to explore, review, and arrange one’s accumulated collection of correspondance “objects”⁴. Csikszentmihalyi and Rochberg-Halton referred to the process of perusing, reviewing, and rearranging one’s possessions as “cultivation”[5], saying that through these activities the meaning of the object is realized; it becomes part of the person⁵.

⁴ When we think of email as object it is not only, or even primarily, the individual emails that encode meaning, for the larger patterns of email use function as meaningful objects - if they are made visible. Here one can see the patterns of daily life, the growth of a circle of acquaintances, the ebb and flow significant topics. It is here that the importance of visualizing the archives is most apparent, for it is often these patterns, rather than the individual emails, that are significant. The exchange of messages with a close friend may be composed primarily of such mundane texts as “hey, free for dinner this evening?” Each message alone may seem unimportant, but their accumulation over time is significant.

⁵ Similar ideas are found throughout the field of material culture studies. Grant McCracken, for example, refers to this behavior as “possession rituals” [24]. His emphasis is more on how new objects transfer their meaning to the owner, but the overall notion is the same.

The cultivation metaphor suggests interfaces that are highly interactive – that allow the user to explore the data, to annotate it, to reconfigure pieces of it. People find exploring interactive visualizations of data far more compelling than they do performing traditional text searches and queries on the same data [34]. Through these explorations, they become more deeply familiar with the material, uncovering new patterns and reflecting on familiar ones – transferring meaning from object to user. (It is useful to note that by designing interfaces in which annotation and learning are integrated with exploring, the process of exploring the data also becomes one of reshaping and refining it – transferring meaning from user to object.)

4.2 Display

Email, as it is currently designed, is a private medium. We do not generally let other people read our email and it is poor etiquette to peer over someone's shoulders to glance at the subject and sender lines in their inbox. Yet as we continue to explore the "email as habitat" metaphor, it is important to consider how visualizations of email could function as a display.

One of the most interesting findings in our initial user studies of *Post-History* and *Social Network Fragments* was that these visualizations became catalysts for personal narrative. Clusters of messages reminded the user of events and names stimulated the recall of stories. While these studies were small and preliminary, they do indicate that it is important to think about email visualizations as a form of display for others. Clearly there is still much work to be done to understand how, when and why people would use email as a display. Here we will simply set out some initial thoughts and questions.

It is important to note here that the visualizations we are discussing show the temporal patterns of emails sent and received and the social networks embedded in the addresses; they do not show the actual contents of the messages. People do, on occasion, display letters, such as a commendation from the head of the company or a grandson's first letter. Our assumption, for now, is that the display of actual messages would follow similar social guidelines.

The closest parallel is the photo-album and thinking about how people display photographs and photo-albums is a good starting point for understanding email as display. We can enjoy looking at our photograph collection alone, but also enjoy showing it to others. Sharing this personal history is a way of revealing aspects of oneself that might otherwise be unknown to the other; it can be a way to establish credibility or status. When we think of looking at a photo-album with others, it is in the context of a personal conversation – similar to the experiences described with *Post-History* and *SNF*. Here, the display is shared, but it is far from public.

There are some very significant differences between email visualizations and photos. One is that photographs are often taken with the purpose of eventually displaying them. They are deliberately created mementos of a time and place. Emails are written for all kinds of reasons, but subsequent display – of the message or the sending patterns – is seldom one of them. Furthermore, photo displays are generally edited; we pick and choose what we want to share. The email visualizations we have discussed thus far are posited on having, ideally, a complete archive to depict. If visualizations are designed for the purpose of display should they have an editing feature, so that users could, for instance, trim their social network to display only the most desirable connections?

Edited or not, there is growing evidence that people are eager to display their social networks. *Friendster*, a web-site launched in early 2003 and growing rapidly, is a place where people post pictures and profiles of themselves – and mutually link to all their friends on the site. One reason it has grown so rapidly is that people are motivated to invite their friends to join so their network will grow. *BuddyZoo*, a site created by a Caltech student this year, is more similar to the social network visualizations we have been discussing. Users upload their AIM buddy lists (their list of friends they chat with online) and the site incorporates it into a growing network and performs a series of analyses for the user, including visualizing the buddy list and measuring popularity ("Popularity ranking: 309586 (percentile: 94.759). 309585 people are more popular, 3972954 are less popular, 143686 are equally popular."). These and other public displays of social networks show that people will indeed display them – and that they create competition. Who has the biggest network? the coolest friends?

Wendy McKay [25] and others have noted that the quantity and type of email received by people in different roles in an organization is markedly different; Duchenaut [7] and Panteli [27] have shown that status and other social structure information can be found in message contents. We can envisage a company in which every employee's work email patterns are visualized and publicly displayed (perhaps on their webpage) – a status symbol for the 21st century. People may become adept at reading the subtle distinctions in such visualizations. Do you know a lot of different types of people? Are they connected to each other – or only through you? The use of email as a display of social status raises the question of reliability and honesty. What is it that people would want to display? How heavily edited is the display? Or would there be a perceived need to archive all mail, deleting nothing, so that the display be as complete as possible? In many work places today, all email must be saved, a rule meant to insure an electronic “paper-trail”. In such places, email has already become a fairly public medium, written and received with the assumption that it is open to review at any point. Indeed, creating a more public display of email might help by reminding people that, at least in this context, email is a public system.

5 Summary and conclusions

Our daily lives are increasingly being lived online. Email is used both at work and at home, by people of all ages, occupations, social and economic classes. Teenagers message their friends throughout the day and night⁶; co-workers use email to collaborate across continents or across the room. It is used to announce births, schedule meetings, arrange blind dates. The email archive contains running conversations, reminders of things to do and of missions accomplished. It is a diary and an address book, a record of momentous events and quotidian detail.

The email programs that exist today may be fine for sending email – the simplicity of the medium is one of its strengths, but they do a poor job of handling the people's growing archives of emails. In this paper we have looked beyond simply fixing the more obvious problems in the current model and proposed instead to visualize this rich source of personal data. We looked at what, exactly, this data is, and went through several examples of potential visualization. We discussed ways that email visualizations could be used that are very different from how email is considered today – as a space for personal reflection and as a display to share with others.

The data in email is rich and complex, and we have just begun to explore how to depict it. How do we show patterns of reciprocity? Differentiate between formal and informal correspondence? What is the significance of these patterns and how do we represent this visually?

Finally, it will be especially interesting to see how such visualizations change the way people use and think about email. Perhaps some will be more diligent about answering messages; others might be more anxious to expand their circle of acquaintances. Hopefully, for all, these visualizations will bring greater understanding and meaning to their correspondence and connections.

6 References

- [1] Baron, Naomi. 1998. Letters by phone or speech by other means: the linguistics of email. *In Language and Communication* **18** p. 133-170
- [2] Boneva, B., Kraut, R. and Frohlich, D. 2001. Using E-mail for personal relationships: the difference gender makes. *American Behavioral Scientist*.
- [3] Ronald S. Burt. 1997. [The Contingent Value of Social Capital](#). *Administrative Science Quarterly*, Vol. 42, No. 2. (Jun., 1997), pp. 339-365.
- [4] Burt, R. S. (2000). [Decay Functions](#). *Social Networks* 22:1-28.

⁶ While this paper focuses on email, many of the ideas and designs are applicable to other media such as instant messaging or online chat. There are some related visualizations that have been created for these media. The history interface in Chat Circles is a depiction of the rhythms of a multi-person real-time conversation; BuddyZoo is a depiction of the extended networks of instant messenger buddies.

- [5] Csikszentmihalyi, M. and Rochberg-Halton, E. 1981. *The Meaning of Things*. Cambridge, UK: Cambridge University Press.
- [6] Donath, J. 1995. Visual Who. In *Proceedings of ACM Multimedia*.
- Donath, J. Viegas, F. and Karahalios, K. 1999. Visualizing Conversation. *Proceedings of HICSS-32*, Maui, HI, January 5-8, 1999.
- [7] Ducheneaut, N. (2002). "[The social impacts of electronic mail in organizations: a case study of electronic power games using communication genres](#)." *Information, Communication, and Society* (iCS), 5(2), pp. 153-188.
- [8] Ducheneaut, N. and Bellotti, V. (2001). "[Email as habitat: an exploration of embedded personal information management](#)." *Interactions* , 8(5), pp. 30-38.
- [9] Dunbar, R. 1996. *Gossip, Grooming and the Evolution of Language*. Cambridge, MA: Harvard University Press.
- [10] Farnham, Shelly. 2003. [Personal Map: Automatically Modeling the User's Online Social Network](#). Position paper for CSCW
- [11] Scott Fertig , Eric Freeman , David Gelernter, [Lifestreams: an alternative to the desktop metaphor](#), *Conference companion to Human factors in computing systems: common ground*, p.410-411, April 13-18, 1996, Vancouver, British Columbia, Canada
- [12] Freeman, E., Gelernter., D., 1996. [LifeStreams: A storage model for personal data](#), *ACM SIGMOD Bulletin* 25, 1, (March 1996), pp. 80-86.
- [13] Freeman, Linton C. 1992. [Filling in the Blanks: A Theory of Cognitive Categories and the Structure of Social Affiliation](#). *Social Psychology Quarterly*, Vol. 55, No. 2, Special Issue: Theoretical Advances in Social Psychology. (Jun., 1992), pp. 118-127.
- [14] Granovetter, M. 1973. The Strength of Weak Ties, *American Journal of Sociology*, **78**, 1360-80.
- [15] Hafner, Katie. 2001. [The 30 year path of email](#). *The New York Times*. Dec. 6 2001.
- [16] Hafner, katie, and Lyon, Matthew. 1996. *When Wizards Stay Up Late*. Simon and Schuster.
- [17] Hafner, Katie. 2003. Putting all your e-mail in one basket. *The New York Times*. June 26, 2003 Circuits. pg
- [18] Halle, David. *Inside Culture: Art and Class in the American Home*. Chicago: The University of Chicago Press, 1993.
- [19] Lars Hallnäs and Johan Redström . 2002. [From Use to Presence: On the Expressions and Aesthetics of Everyday Computational Things](#). In *ACM Transactions on Computer-Human Interaction (ToCHI)*, Vol. 9, No. 2, June 2002, pp. 106-124. ACM Press.
- [20] Hill, W., & Hollan, J. (1993). History-Enriched Digital Objects. In *Proceedings of Computers, Freedom and Privacy* (CFP 93)
- [21] William C. Hill , James D. Hollan , Dave Wroblewski , Tim McCandless, [Edit wear and read wear](#), *Proceedings of the SIGCHI conference on Human factors in computing systems*, p.3-9, May 03-07, 1992, Monterey, California, United States
- [22] Hollan, James D. Personal communication
- [23] M. Lansdale, The psychology of personal information management, *Applied Ergonomics*, Volume 19, Issue 1, March 1988, Pages 55-66.
- [24] McCracken, Grant. 1990. *Culture and Consumption*. Bloomington, ID: Indiana University Press
- [25] MacKay, W. 1988 More than Just a Communication System: Diversity in the Use of Electronic Mail. In *Proceedings of ACM CSCW '88: Conference on Computer-Supported Cooperative Work*. Portland, Oregon: ACM.

- [26] Mock, Kenrick (2001). "[An Experimental Framework for Email Categorization](#) and Management." 24th Annual ACM International Conference on Research and Development in Information Retrieval. September 9-13, 2001. New Orleans, LA.
- [27] Panteli, Niki. 2002. Richness, power cues and email text. In *Information and Management* **40**, pp 75-86.
- [28] Pew Internet Project. 2000. The holidays online: Emails and e-greetings outpace e-commerce. Report no. 29, Dec. 31, 2000.
- [29] Pew Internet Project. 2002. The internet goes to college. Report no. 71. Sept 15, 2002.
- [30] Quan, Dennis; Bakshi, Karun; Huynh, David; and Karger, David R. 2003. [User interfaces for supporting multiple categorization](#). In *Proceedings of Interact 2003*.
- [31] Reddy, Madhu C. and Paul Dourish. 2002. [A Finger on the Pulse: Temporal Rhythms and Information Seeking in Medical Work](#). In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work CSCW 2002* (New Orleans, LO). New York: ACM.
- [32] Roberts, W. (1989). [Affiliation Structures in Groups of Young Children: A Computer Simulation](#). *Developmental Psychology*, 25, 805-811.
- [33] Warren Sack, "[Discourse Diagrams: Interface Design for Very Large Scale Conversations](#)" in the *Proceedings of the Hawaii International Conference on System Sciences, Persistent Conversations Track* (Maui, HI: IEEE Computer Society, January 2000).
- [34] Shneiderman, B. (1994). *Dynamic queries for visual information seeking*. IEEE Software, 11, (6), 70-77.
- [35] Spertus, E. Smokey: Automatic Recognition of Hostile Messages. *Proceedings of the 14th National Conference on AI and the 9th Conference on Innovative Applications of AI*, pp. 1058-1065, Providence, Rhode Island, 1997.
- [36] Tyler, Joshua R., Dennis M. Wilkinson and Bernardo A. Huberman. nd. "[Email as Spectroscopy: Automated Discovery of Community Structure within Organizations](#)"
- [37] Tyler, Joshua R. and Tang, John C. nd. [When can I expect an email response? A study of rhythms in email usage](#).
- [38] Venolia, Gina and Neustaedter, Carman. 2003. [Understanding sequence and reply relationships within email conversations: A mixed-model visualization](#). Microsoft Research Technical Report MSR-TR-2002-102.
- [39] Viegas, F., Boyd, D., Nguyen, D., Potter, J. and Donath, J. 2004. Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. To appear in *HICSS 2004*.
- [40] Watts, Duncan. 2003. *Six Degrees*. New York: W.W. Norton.
- [41] Whittaker, S., Jones, Q., and Terveen, L. (in press). [Contact Management: Identifying Contacts to Support Long Term Communication](#). To appear in *Proceedings of Conference on Computer Supported Cooperative Work*. New York: ACM Press.
- [42] Whittaker, S., Jones, Q., and Terveen, L. (2002). [Managing Long Term Conversations: Conversation and Contact Management](#). In *HICSS 2002*.
- [43] Whittaker, S. and Sidner, C. (1996). [Email overload: exploring personal information management of email](#). In *Proceedings of CHI'96 Conference on Computer Human Interaction*, NY: ACM Press, 276-2
- [44] Yates, J., Wanda J. Orlikowski, and Kazuo Okamura. [Explicit and Implicit Structuring of Genres: Electronic Communication in a Japanese R&D Organization](#), Working Paper Number: 3809, April 1, 1995; *Administrative Science Quarterly*, 39, 4, pp541-574, December 1994
- [45] Yiu, Kelvin S., Baecker, Ronald; Silver, Nancy, and Long, Byron. 1997. [A Time-based interface for electronic mail and task management](#). In *Design of Computing Systems: Proceedings of HCI International '97*. Vol. 2. Elsevier, 19-22.

[46] Zhang, J. and Mostafa, J. 2002. Information Retrieval by Semantic Analysis and Visualization of the Concept Space of D-Lib® Magazine. In *D-Lib Magazine*. Volume 8 Number 10. October 2002.

Figures

Figure 1 *Social Network Fragments*: The above image shows a temporal slice of a full network. Boxes on the right represent 2 week periods – selecting one shows the user’s network during that time.

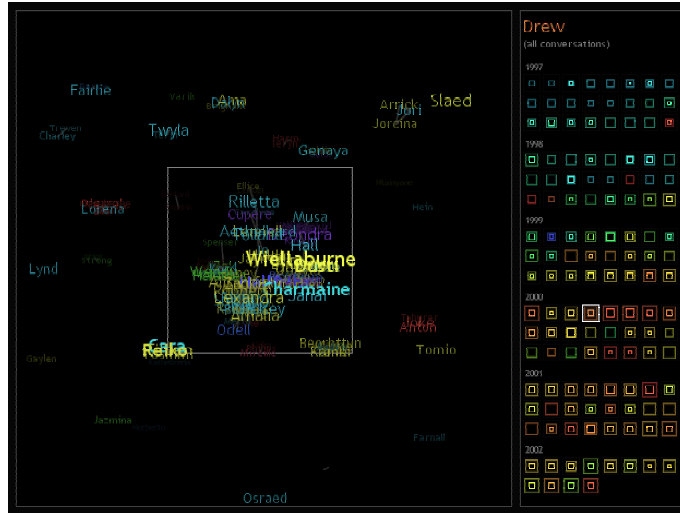


Figure 2 *Post history*. This image shows the calendar on the left and the set of names the user has corresponded with on the right. One correspondent has been highlighted and yellow boxes indicate messages exchanged with that person.

