

Visualizing a Personal Social Network of Email Archives For Re-Finding

Xiaoyan Yu, Pengbo Liu, Mohammad Alkandari, and Manuel A. Pérez-Quiñones

Department of Computer Science

Virginia Tech

Blacksburg, VA 24060

{xiaoyany, pengbo, mkandari, perez}@vt.edu

ABSTRACT

Email has become an important communication medium that people routinely rely on to achieve their daily work. However, managing email accounts for life-long communication is not an easy task. Users have different strategies for archiving their messages and have varied understanding and expectations of retrieval and how to use that capability. We have designed and implemented a tool, VisPEAM that applies the benefits of visualization of social network in personal email archive management. This helps users re-find email messages by using people's name, folder and subjects as search criteria. The graphical representation helps users to find their old emails by exploring the social relationships in different folders as well as shows how frequently those people exchange email messages regarding a particular topic. Moreover, the graphic visualization helps people manage their growing collections of email messages with respect to their social relationships.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Management, Design, Human Factors

Keywords

Email management, social network, visualization.

1. INTRODUCTION

In societies where the technology is readily available, email has become a universal communication medium in people's daily life. Their email archives reflect their social lives and also the stories of their happiness and sadness while current email clients hide that social information. The high volume of email messages and unsuccessful filing strategies may hinder people to re-find email from their personal archives. Bush in 1945 envisioned a personal information management tool, MEMEX. It could not only keep all the information but also make good use of them [2]. However, current personal email archive management tools are more successful in supporting email storage than in later re-use. In

addition, instead of relying on keyword search to get the desirable information, people tend to use small, local steps and context search cues especially, time and other people's names to re-find information in email [3]. We developed VisPEAM, an add-on feature to email clients. It visualizes social network on personal email archives to uncover the social flavor of the architecture. It also arranges the search results in the context of social network to shorten the steps in building a re-finding chain.

The paper is organized as follows. We will first introduce related work, followed by a brief description of personal email archives. We develop scenarios to illustrate the benefits of our solution then focus on how our tool is designed and implemented. Finally we conclude our work and point out future work.

2. RELATED WORK

Many studies focus on issues relevant to email archive management such as email overloading and re-finding. For example, Whittaker et al. [1] proposed technical solutions for the email overload problem, in which they presented qualitative and quantitative information about the use of email for task management, personal archiving, and asynchronous communications. The research of Gwizdka [4] seeks to gain understanding of individual differences in email behavior. This study presented results from a questionnaire-based study that discusses how users manage their email messages and how they keep the important messages for future retrieval.

Implementing visualization in email tools is highly recommended [5] as a way to help users understand the relationships between the various messages as well as the relationships between different people.. For example, Venolia and Neustaedter [6] described thread structure of a person's email and how it is evolved along time.[7] Mansmann et al. [8] applied Self-Organizing Maps (SOM) to visualize archival emails with the feature of important terms in the Subject field. Visualization of social networks has a good history in which many people have been interested in applying social network in email archive management, especially in conjunction with other features such as time. Vidas [9] discussed important concepts in which, both Post History and Social Network Fragments (SNF) focus on two major dimensions of email archive: people and time. Their visualization provides striking insights into the experience of social connectedness over time. These insights basically enable users to better manage how they invest their time and energy into personal relationships as well as improve the whole sense of well-being. Our visualization is very like SNF but more focuses on facilitating re-finding email.

Some studies also focus on visualization of social network on areas other than email. Vizster [10] is designed and implemented a visualization system for playful end-user exploration and navigation

of large scale online social networks. Our implementation of VisPEAM is based on Vizster, built on the Prefuse package [12].

3. PERSONAL EMAIL ARCHIVE

Whittaker [1] observed three different email filing strategies: *no-filers* who do not make use of any folders and keep a huge incoming email box; *frequent-filers* who attempt to file daily and have folder structured archives; and *spring-cleaners* who occasionally file and usually have an extensive folder structure. People with different filing strategies have different email archive structures. We therefore also take into account, folder information (if it is available) when analyzing relationship in an archive. Our visualization tool is tested on Enron's corpus. The Enron corpus was made public after legal investigation. After removing duplicated messages, there are a total of 200,399 messages belonging to 158 users. [11] We selected a single person Sally's email. In her email archive, all email messages have been organized in predefined folders.

3.1 Folder Relationship

A natural grouping of people is through the use of folder

information. Each folder contains a number of messages usually relating to the same subject. Therefore, we consider people who appear in the same folder to be connected by the same topic. If there are a great number of common names appearing in two folders, the folders have a high degree of correlation.

3.2 People Relationship

To extract people's relationship from email folders, we take advantages of email header information, as all people involved in email threads appear in either the From, the To, or the Cc fields. Between two people, say Person A and Person B, who appear in an email header, there can be five types of relationships: FromTo(A sends email to B or A receives email from B), FromCc(A carbon copies email to B or A receives a copy from B), ToTo(A and B receives the same email), CcCc(A and B receive the same copy), ToCc(A receives a email message and B receives a copy or vice versa). These five types of relationships play different roles in depicting people's closeness. For example, a name which appears in the To field of an email has a stronger relationship with the name in From field than a name in the Cc field.

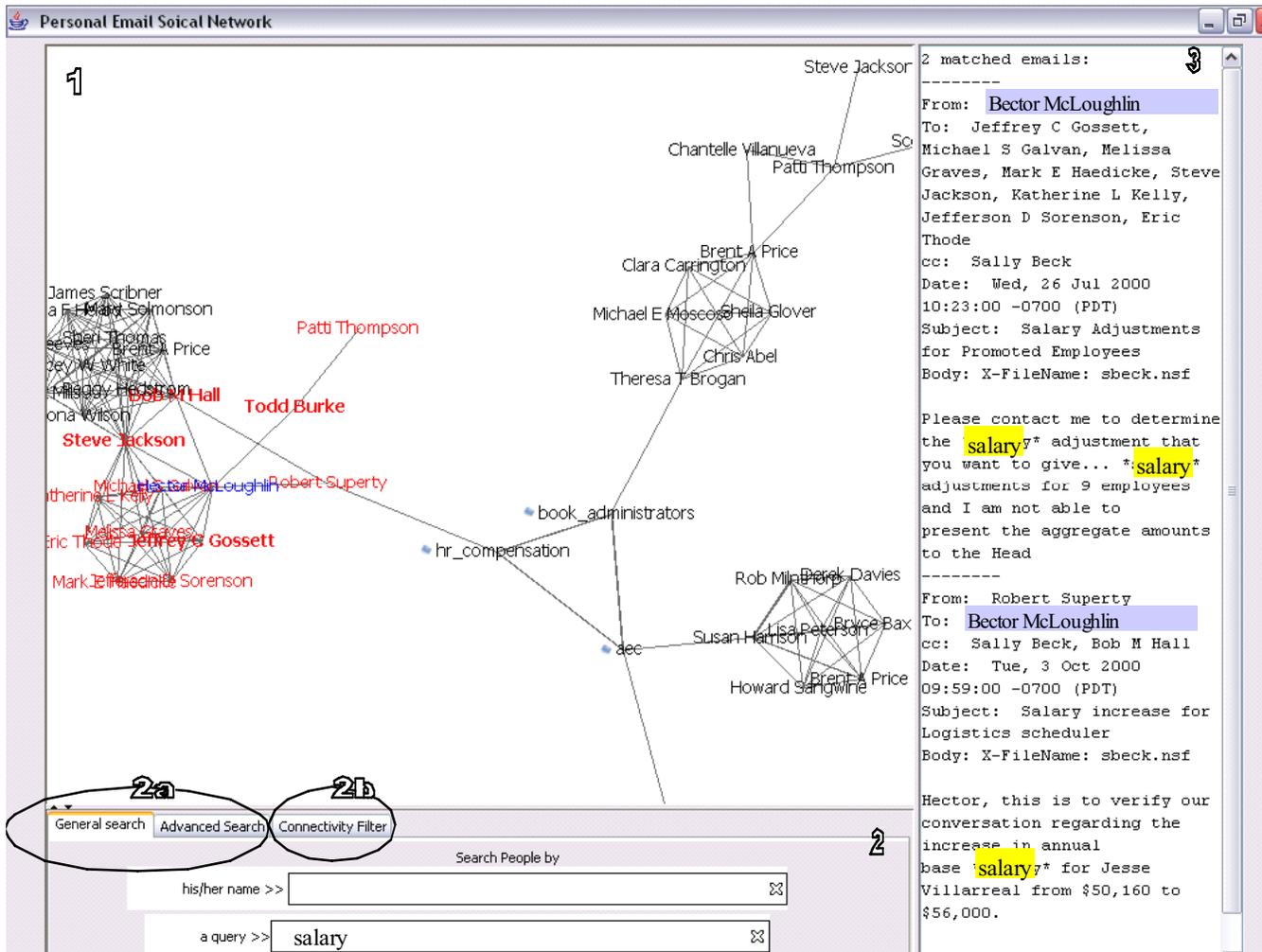


Figure 1. VisPEAM tool: 1) view panel; 2) control panel with 2a) search filter with search for 'salary' and 2b) connectivity filter; 3) result panel shown two matched emails for the selected correspondent 'BectorMcLoughlin'.

4. SCENARIOS

Before describing our tool, we will illustrate some scenarios to explain the benefits email archive owners such as Sally can get from a tool like ours.

Scenario 1: Sally might discover unexpected persons with the help of the interactive, visualized social network. When exchanging emails, a sender might carbon copy to some related people that Sally has not realized. The visualized networks such as those shown in Figure 1 can give Sally an idea of other people involved in the email. Highlighting neighbors as shown in Figure 2, one of our tool’s features, would also help Sally, if she were interested in those related people.

Scenario 2: Sally thinks that she has filed a particular email in the folder ‘book_administrators’ but does not find it there. By using our tool, she can explore the neighboring folders to see whether that email is in those folders or not. In addition, she might get to remember the sender, Brent A. Price, of email during the exploration. When she searches for ‘Brent’, the name would appear in different places and she then can click on the best fit one (for example, the one resides in the ‘aec’ folder shown in Figure 2 which it is more possible in her mind to file the email) to fast locate that email.

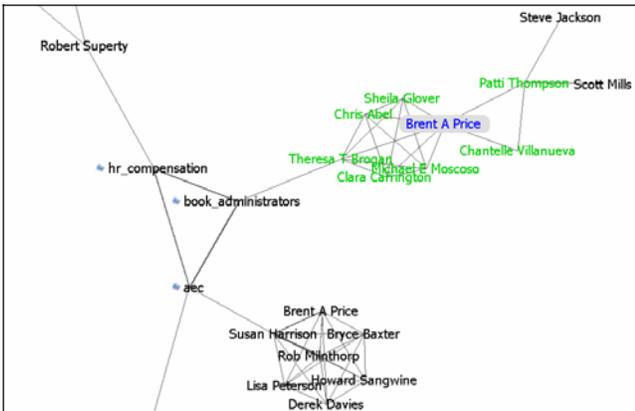


Figure 2. Layout of personal social network with ‘Brent A Price’ focused on.

Scenario 3: Sally might not remember who has sent a message, say relevant to a meeting. She roughly recalls the topic. Using this topic, our tool will highlight relevant people shown in Figure 1 in red. This might help her recall the sender of this message. If there are too many people highlighted to help recall, she might select people most likely to be involved in that message. For example, the center of a communication community, such as Brent A. Price in the ‘book_administrators’ folder who communicates with most people in that folder shown in Figure 2, might be a good choice.

5. VISPEAM

Our tool, Visualization for Personal Email Archive Management (VISPEAM), has as a goal to help an email archive owner explore his/her email archive and re-find desired information. We apply visualization to a derived social network and integrate a search functionality. The tool consists of 3 components: view panel, control panel, and result panel, shown in Figure 1. The view panel renders a personal social network. The filters on the control panel can highlight and hide some portions of the network. Two filters are

supported: search filter and connectivity filter. The result panel shows the search results.

5.1 Visual Items

A personal social network consists of nodes and edges. The characteristics of the network are summarized in Table 1. Correspondents and email folders are presented as nodes. The same correspondent in different folders is presented as different nodes. The attributes of a correspondent node include the person’s name and the closeness (explained in the section 3.2) with the owner. A correspondent node is rendered as a label with this person’s name in the text field. The archive owner is not put into the network at all to reduce trivial edges since everyone is connected to the owner. Instead, the text font of a correspondent node is bold or regular to present their closeness to the owner. The attributes of a folder node include the folder’s name, a folder image to distinguish it with a correspondent node. A folder node is rendered as a label with the folder name in the text field and the image in the image field. The ‘book_administrators’ node in Figure 2 is a folder node. An edge connecting two correspondent nodes represents the existence of relationship between the two people. The closer these two people are, the thicker the edge is. An edge between two folder nodes presents common correspondents in these two folders. The width of such an edge also represents the relative number of the common correspondents residing in the two folders. There are also edges between folder nodes and some of correspondent nodes to visualize correspondents in different folder. For example, all the correspondents in the clique shown at the bottom of Figure 2 have emails in the folder ‘aec’.

Table 1. Summary of visualized items.

Items	Types	Attributes	Representation
Node	Correspondent	Person’s name	Text field
		Closeness	Text font
	Folder	Folder’s name	Text field
		Folder distinguisher	Image
Edge	Correspondent-correspondent	Email traffics	Edge width
	Folder-folder	Common correspondents	Edge width
	Correspondent-folder	Whose email filed in which folder	A very light edge

5.2 Layout

The network is arranged in the force-directed layout, where each node repels each other and links act as springs. People with more links among them tend to cluster together and naturally form a community. People within the same folder are closer to each other so it is easily to recognize who are in a particular folder. Each community organized by folders has its own pattern. Some of them are cliques such as the ‘aec’ folder in Figure 2. Some contain several sub-communities with articulate correspondent nodes such as the ‘book_administrators’ folder in Figure 2. These patterns would help the owner dig out information from the archive, as illustrated in the third scenario in the section 4.

5.3 Interaction

A user may be overwhelmed upon facing a complex social network even when depicting his/her own email archive. We hence provide plenty of interactions on the network for users to explore their archives. Our tool supports basic mouse interactions such as panning and zooming. Mouse across a node fills its text field into grey color and changes the text color from black to blue. In this way, as shown in Figure 2, the name is much clearer to recognize than the one without fill-in color. Single left-click on a correspondent node is to select the node for further interaction. Double left-click on a person's name will show his/her email excerpts matching search criteria. Double left-click on an edge between correspondent nodes will show excerpts from the emails exchanged.

The owner might be interested in whom a correspondent communicates with besides himself/herself (see the first scenario in the section 4). The owner can follow each out-link from the person to find all the person's neighbors if it is easy to follow or zoom in the region around the person. We also support an easier way, neighbor highlighter, to help the task. When the mouse is put across a node, not only the node itself will change but also its neighbors will change from black to green. In Figure 2, it is easy to recognize that in the 'book_administrator' folder Brent A. Price communicates with everyone but only 2 people Steve Jackson and Scott Mills.

Our tool supports 3 search filters: name filter, content filter, advanced filter (content plus date). The name filter highlights nodes whose text field prefix match the searching names. The owner would get benefits from the prefix search on text fields without the necessity to complete the names they are looking for. The name filter also makes the owner aware of which folders a particular correspondent's emails reside. The content filter allows the owner to search his/her archive by email subject or body and highlights the nodes of people who are involved in such emails by changing the color from black to red. Figure 1 shows the results of applying the content filter. The advanced filter can narrow down the search on the arrival date of email messages. Upon applying the content filter or the advanced filter, excerpts of emails associated with highlighted people are shown on the result panel by double clicking any of these nodes. An excerpt includes email headers and sentences around highlighted query terms in the email body. Currently we highlight the query terms by adding "*" around them instead of coloring them (see the right panel of Figure 1 for examples).

Our tool runs on Java 2 Standard Edition, version 1.5. The user interface and visualization are built on the Prefuse package [12] which offers many useful visualization and interface components. After analyzing correspondent and folder relationship from email archive, a GraphML [13] file describing the social network is generated to feed as data to the visualization. The name filter is implemented by in-memory search. The content and advanced filters are supported by a search using the Lucene package [14].

6. CONCLUSION AND FUTURE WORK

We have developed a tool of visualizing the social network of large volumes of personal email messages. The benefit of building up a visualized social network lies in its potential ability to reduce the effort of re-finding email messages. This tool is designed to help people recall necessary information by exploring the social network. For example, a user could retrieve messages even if the

original sender was forgotten, or discover new relationships by exploring the social network. The proposed tool has a simple user interface to facilitate the re-finding process. This interactive user interface can highlight the same person in different folders, change focus on the social network, and zoom-in on individual folders.

In the future, we plan to conduct a usability study to verify the effectiveness of our tool by observing how users interact with the tool given the above scenarios. We expect to gain further insight into the utility of this tool based on that study.

7. ACKNOWLEDGMENTS

We thank Travis Rose and Uma Murthy for their comments on this paper. We would also like to thank the students in the CS 5984 Personal Information Management course at Virginia Tech for the insightful in-class and forum discussions.

8. REFERENCES

- [1] Whittaker, S., and Sidner, C., *Email overload: exploring personal information management of email*. Proceedings of the SIGCHI conference on Human factors in computing systems: common ground, ACM Press, 1996: p. Pages: 276 - 283
- [2] Bush, V., *As we may think*, in *The Atlantic Monthly*. 1945. p. 101-108.
- [3] Cutrell, E., S.T. Dumais, and J. Teevan, *Searching to eliminate personal information management*. Commun. ACM, 2006. 49(1): p. 58-64.
- [4] Gwizdka, J., *Email task management styles: the cleaners and the keepers*. CHI '04: CHI '04 extended abstracts on Human factors in computing systems, ACM Press., 2004: p. Pages: 1235 -1238.
- [5] Somervell, J., McCrickard, D., North, C., and Shukla, M., *An evaluation of information visualization in attention-limited environments*. VISSYM '02: Proceedings of the symposium on Data Visualisation 2002., 2002: p. Pages: 211 - 216.
- [6] Venolia, G., and Neustaedter, C., *Understanding sequence and reply relationships within email conversations: a mixed-model visualization*. CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press., 2003: p. Pages: 361 - 368.
- [7] Perer, A., B. Shneiderman, and D.W. Oard, *Using Rhythms of Relationships to Understand Email Archives*. To appear in the Journal of the American Society for Information Science and Technology (JASIST), 2006.
- [8] Mansmann, F., T. Schreck, and D. Keim. *MailSOM - Exploration of Electronic Mail Archives Using Self-Organizing Maps*. in *CEAS*. 2005.
- [9] Viégas, F.B., et al. *Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments*. in *Proceedings of the 37th Hawaii International Conference on System Sciences*. 2004.
- [10] Heer, J., and Boyd, D., *Vizster: Visualizing Online Social Networks*. Computer Science Division and School of Information and Systems - University of California, Berkeley., 2005.
- [11] Klimt, B. and Y. Yang. *Introducing the Enron Corpus in CEAS*. 2004.
- [12] Prefuse, <http://prefuse.org/>.
- [13] GraphML, <http://graphml.graphdrawing.org/>.
- [14] Lucene, <http://lucene.apache.org/>.