Email Visualization

Anbang Xu

In this paper, we propose two layouts, email wheel and thread flow map, to visualize the email dataset. Email wheel would reveal different relational groupings from the email dataset, where spam mail can be highlighted. The flow map thread can visualize the different phases of the email thread and help users to trace the history of the email thread.

Previous work

The tools for interactions with email collections can be divided into six categories [1] (see Table 1). As the purposes of interaction with email dataset are different, the email visualization methods are also different. The focuses of our work is designing a layout to help users explore an archive of an organization's messages.

Table 1. Types of email interactions.

<table>
<thead>
<tr>
<th></th>
<th>Individual</th>
<th>Organizational</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>Managing an individual user's current inbox</td>
<td>Managing current email within an organization</td>
<td>Managing current conversations within a public space</td>
</tr>
<tr>
<td>Archived</td>
<td>Exploring an archive of an individual's messages</td>
<td>Exploring an archive of an organization's messages</td>
<td>Exploring an archive of a public space</td>
</tr>
</tbody>
</table>

Enron Corpus Viewer [2], a force mode based layout, is proposed to visualize an archive of an organization's emails (see Figure 1 (a)). However, the irregular locations of nodes may make users to perceive the information from the layout inefficiently. The “My Map” [3] is a circular layout, focusing on visualizing an archive of an individual's emails (see Figure 1 (b)). “My Map” allows the user to explore the relationship between him and all other people, and the changing trend of the relationship. However, both of “My Map” and Enron Corpus Viewer have the problem that when the number of nodes in the layout increases, the edge crossings would become excessive.
Our work

Email wheel

The basic idea of this layout is inspired from “My Map”. Our contribution is reducing visual cluster by bundling edges together and using color to encode the “worth” of email.

Recently, Holten et al. [4, 5] propose a new graph visualization technique—edge bundling to visualize the hierarchical data (Figure 2). This method can also be extended to visualize nonhierarchical data such as email data.

How much are the email worth? We can define a metric to evaluate the worth of email. For example, if the email is important to the user, he would reply this email in a short time or click the email several times to read it repeatedly.

Using this layout, the user is firstly required to select a time range he wants to explore. Then the email wheel would visualize the email dataset in the time range. Based on the cluster of the edges, different relational groupings would be clearly revealed (see...
In addition, the user can also select a series of periods of time. Each period of time would be represented as an email wheel. He can compare the different email wheels and find the temporal ebbs and flows in various relationships.

Figure 3. In addition, the user can also select a series of periods of time. Each period of time would be represented as an email wheel. He can compare the different email wheels and find the temporal ebbs and flows in various relationships.

Figure 3. The layout of email wheel. The node represents the person and the link represents email between two persons. The color encodes the worth of email. The green color represents the most useful email, which means the email is replied by the user very quickly or read by the user repeatedly. The red color represents useless email, which means spam or the user ignore this email and never reply it.

Thread flow map

People often received a lot of emails on one thread. In the current email services, the email of one thread can only be kept straight. However, the topic of this thread may change over time. One topic can be separated into different topics when people keep discussing one problem. Also, two topics can be merged into one topic. For example, when the students in one class discuss the spring break activities, firstly, all of the students decide to barbecue. Then some of them begin to discuss when to barbecue and others talk about where to barbecue. After a while, the topic about where to barbecue separates into two topics: the weather and price of food. When the students talk about the weather, they find that it would be rainy on the morning, so they restart to discuss when to barbecue and set up the final time schedule. Figure 4 show the history of the thread.
Basically, text mining method will be applied to analyze the content of emails and the proposed layout will visualize the results. We firstly extract features such as key words from the email. Then the emails can be clustered into several groups. In this layout, each node represents one group of emails (see Figure 4).

![Figure 4. The thread flow map. The node represents the topic of the thread. The size of node and the width of edge encode the number of people involved in this topic. The color encodes the number of emails (one person can send more than one emails on the same topic). It means that even if the size of two nodes is same, the color of the nodes can be different (the more emails involved, the more red coloring it show). The timeline is drawn from left to right.]

**Conclusion**

Two layouts -- email wheel and thread flow map are presented to visualize an archive of an organization's emails. Email wheel can real different relational groupings from the email dataset. Our contribution is reducing visual cluster by bundling edges together and using color to encode the “worth” of email. Thread flow map can help user to perceive the changing trend of topics in the thread.

**Reference**


