

# Who have I been talking to?

Tiago Garcia  
Instituto Superior Técnico  
Lisbon, Portugal  
tiago.garcia@ist.utl.pt

João Aires  
Instituto Superior Técnico  
Lisbon, Portugal  
joao.aires@ist.utl.pt

## 1. INTRODUCTION

Nowadays, we can say that the computer is used a lot for communication by many people around the globe. From social networks to electronic mail, billions of people use computer to communicate and interact with others. In social networks we got for example Facebook, with more than 600 millions users<sup>1</sup>, while email is a technology used even more. Actually, in current times, the use of email is something close to irrational and numbers prove that people consider it an indispensable tool in their lives for either business or pleasure. According to numbers released by Royal Pingdom concerning the 2010 use of email<sup>2</sup> already excluding the 89% considered spam, there were over 32 billion email messages sent every day. It is estimated that the number of email users is about 1,88 billion around the world. With a few calculations is easy to understand that the average person gets around 17 email messages everyday, which makes about 6000 messages in the whole year. This number is an average, so there are a lot of people that probably get much more than that.

So, how can we keep track of our electronic relationships? In this paper, we try to propose a way to visualize information in order to keep track of it. We chose to use email in this work and the reasons were presented on the previous paragraph. The use of email is much more massive than any other platform and email is much more ubiquitous than any other way of communication possible via computer and internet.

The main objective of our work is to answer the question "who have I been talking to?". In addition to that, we propose a solution that also allows the user to understand the "why" of the answers found for that question as well as understand the evolution of the relationships we maintain by email as time goes by.

The structure of this paper consists of basically, after the introduction, a section about related work where a little bit of state-of-the-art in this area is presented, followed by some sketches of the visualization and the tasks the system is supposed to support in the end.

## 2. RELATED WORK

<sup>1</sup>Facebook have 600 million users according to [http://www.msnbc.msn.com/id/40929239/ns/technology\\_and-science-tech\\_and\\_gadgets/](http://www.msnbc.msn.com/id/40929239/ns/technology_and-science-tech_and_gadgets/)

<sup>2</sup>Internet in numbers at <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers/>

## PostHistory and Social Network Fragments

Throughout our lives, we humans tend to associate objects to events and experiences thus assigning a personal and, often, inestimable value to those objects. The fact that our life is increasingly digitized does not alter this need, creates however new challenges to get such social artifacts. *PostHistory* and *Social Network Fragments*[4] are visualization tools that show the users social digital artifacts through the visualization of high level patterns of their email habits: *PostHistory* focuses on direct interactions between users with each of their email contacts ; *Social Network Fragments* allows users to visualize the social network patterns that emerge due to email exchange. Somehow like photos, these visualizations are social artifacts for remembering and storytelling providing a tangible link to digital interactions letting users reflect on their patterns.



Figure 1: The *PostHistory* interface

As shown in figure 1, the PostHistory interface is divided into two panels: on the left, the calendar panel and, on the right, the contact panel. The first one shows the intensity of email exchanges over time and the second one, the names of the people with whom the user has exchanged email. On the calendar panel, each square represents a single day and its size it's proportional to the quantity of email received on that day. Also, the brighter the color of a given square, the more directed to the user were the emails of that day (for instance, an email sent only to the user is "highly directed" and an email sent to a mailing lists is "not directed at all"). Regarding the contacts, there are three possible representations: vertical, circular and alphabetically. The user can

also click on a person name on the contact panel to see yellow squares displayed on top of each day that the person has sent an email, or click on a specific day on the calendar to get the people names highlighted. There's also an animation that simulates the passage of time.



**Figure 2: The Social Network Fragments interface**

*Social Network Fragments* is an interactive animation showing data evolving over time. Its interface is also divided into two panels (see figure 2): the primary social network panel and a history panel. The history panel describes each time slice with two squares. The outer square indicates the number of connections that occurred during the period of time and inner square indicates the number of knowledge ties. During the animation, the current time slice is highlighted. The user can click on a time square to force the application to continue the animation from that point. The network panel shows the network, but only people actively communicating with the user during a given time slice are shown in this view. Names with larger font represent individuals strongly tied to others and to the user. The user can click on an individual's name to see just that person's frequency of connections over time slices. It's also possible to do zoom-in to certain regions of the network.

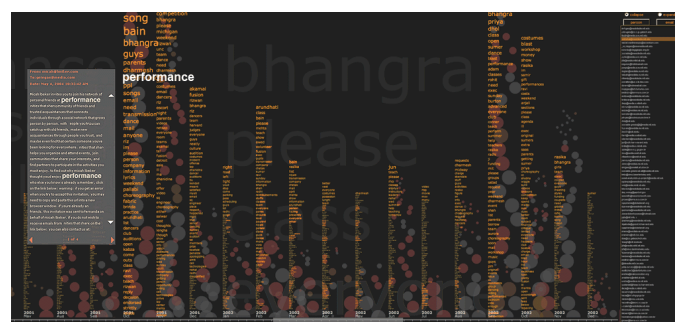
**Strengths:** Interactive and, experiments show, quite effective and exciting to use.

**Weakness:** Too high level: users have no access to the actual content of the exchanged messages what can be frustrating because users miss the all context. Besides that, it's also not possible to follow the exchanged messages with another person over time so we don't really have an overall view of who are contacting to over time.

## Themail

The article's authors propose *Themail* as an application to allow the visualization of the contents of a person's email archive. The main goal of this application is to allow a visualization that portrays relationships among individuals using the exchanged messages preserved in email archives. What the authors claim is that with the increasing number of interactions occurring via email, this information has become a "valuable records of people's relationships".[5] Unlike other email visualizations that only look at header information, *Themail* relies on the content of the messages to build a

visual display of interactions between individuals over time providing clues about the topics individuals discuss. With such goal in mind, the application uses an algorithm to calculate the topic words which are then arranged in columns (a column corresponding to a month or year) along a timeline. Large faint words on the background represent the most used terms by the user with a friend over an year and yellow words on the foreground over a month (see figure 3). The selection and size of the words to be displayed are based on the frequency and on how distinctive the word is to this specific relationship when comparing to the rest of the archive. The colored circles represent email messages exchanged: the circle size stands for the length of the message and the color for the direction of the message (i.e. incoming or outgoing). Figure 3 illustrates how someone can get the messages associated with a particular word just by clicking on that word (on this case, the word "performance"). The email messages that contains such word will then appear in an information box (figure 3, on the left)



**Figure 3: Themail interface**

These tools give us the ability to follow our email messages history and understand, following the timeline, which topics were most discussed throughout this time and watch the underlying trend. The authors also asserted that this application released the storyteller that's inside each of us and that, in many cases, users recall events and interactions of which they simply no longer remembered. It is like looking through a photo album, some users said. This paper also reveals that the two modes of using this system (which the authors called the *haystack* and the *needle* mode of interaction) had a substantial number of usage patterns emerge.

**Strengths:** The possibility of reviewing our past through our email archives reveals to be quite interesting and the way this story is shown is truly faultless in our opinion. Using *Themail* we can easily watch the evolution of our relationships with each one of our contacts.

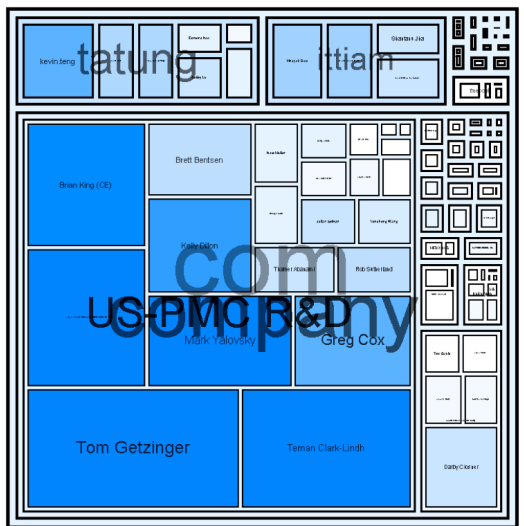
**Weakness:** It's not possible to correlate the evolution of our relationships with different people. We also feel that it's also lacking some kind of general view that would contemplate all our relationships and then compare our individual relationships with it.

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

This article[2] presents three different visualizations of our email archives letting us know more about the people we

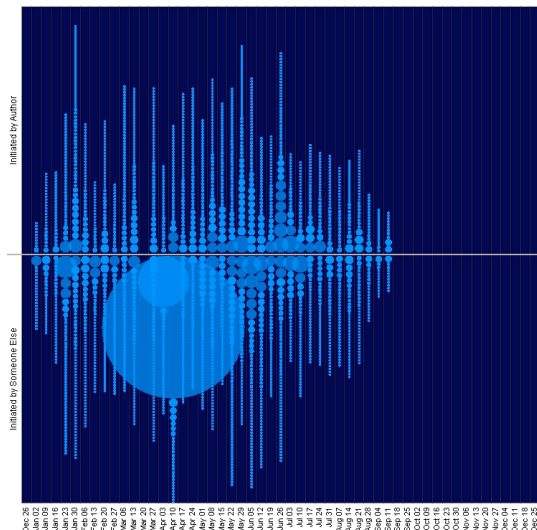
talk to.

The first visualization (*Correspondent Treemaps 4*), based on treemaps, shows a set of boxes each one of them grouping email contacts from the same domain. The size of the boxes is proportional to the number of received emails and its tone reflects the number of unread emails from that domain. It's possible to interact with these boxes in order to read the corresponding emails or even entering on the next view.



**Figure 4: An example of a Correspondent Treemaps**

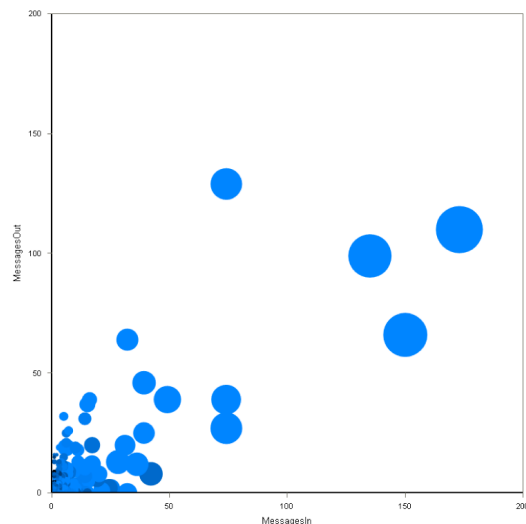
The next view, *Author Lines 5* is just like an horizontal histogram that captures the activity patterns in terms of initiation of new conversations (represented by the bubbles above the horizontal axis) and replies to threads initiated by others (the bubbles below the horizontal axis). The bubbles size is proportional to the number of messages on that week.



**Figure 5: Author Lines**

The last visualization is *Correspondent Crowd 6* a scatter plot generated based on the number of messages sent to

the correspondent against the number of messages received from the correspondent. Here, each correspondent is represented as a circle whose size is a reflection of the total number of messages received from that correspondent in the selected time period. The color represents how long as the correspondent sent his/her last message as the saturation decrease over time.



**Figure 6: An example of a Correspondent Crowd**

**Strengths:** The presented visualization complement each other by showing important details about the people we contact by email on a daily basis.

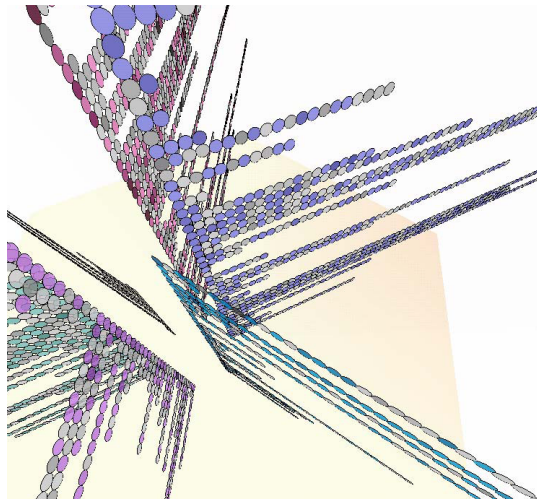
**Weakness:** Our relationships to each one of our contacts is not well explored and the evolution of such relationships is not explored at all.

## CrystalChat

While other social visualizations often focus themselves on social networks, "CrystalChat focuses on the social interaction centered around one person, supporting selfexploration of one's own chat history"[3] revealing the patterns associated with an individual.

This goal is achieved by incorporation the text messages in a 3D structure that integrates an individual's chat interaction pattern with a temporal trace of history, creating an interactive visualization. The *CrystalChat* structure (figure 7) consists of several facets. Facets contain all the conversations, messages, etc, between the user and a given friend. Each circle corresponds to a exchanged message and its color indicates who sent the message (gray for the user, other color for the friend). Each row of circles is a conversation and the bottom conversation is the oldest one. Using this structure is possible to visualize the several conversations we already had and find out, at a glance, very interesting aspects of our pattern of conversation: how many conversations we initiated, what are the people with whom we talk the most, etc.

**Strengths:** The information provided, although it is just about chat messages, is quite complete and is organized in



**Figure 7: A top level view of the CrystalChat 3d structure**

a very interesting way providing a fairly large palette of information about user's conversation pattern.

**Weakness:** The possibility of the 3D structure of accommodating information is directly proportional to the complexity of navigating and manipulating it. Therefore, as it was expected, the use of the 3D structure can be a little challenging and/or frustrating. It's also not that easy to get the big picture.

### 3. VISUALIZATION SKETCHES

In this section, sketches of the proposed solution are presented, in order to explain more about the visualization we want to develop to achieve goals.

In figure 8 we present an image of the general visualization. It has three views. The first and main one shows the amount of emails sent by the user's contacts in each month. A line divides the view horizontally, creating a division between received mail (top) and sent mail (bottom). Each month division has two bars (normal on the top half and inverted on the bottom half) with different partitions, also having each partition its color. To understand the partitions the reader should picture a bar chart ordered from tallest to shortest. Now, if that chart was 3D, the normal view could be considered a front view. What we see in the image is like the side view, with the smallest in front and the tallest on the back. Each bar would represent a different contact with the size being directly proportional to the number of emails sent or received. Now, considering the visualization presented, the partition on top represents the contact that in that time sent (top half) or received (bottom half) more emails to or from the user.

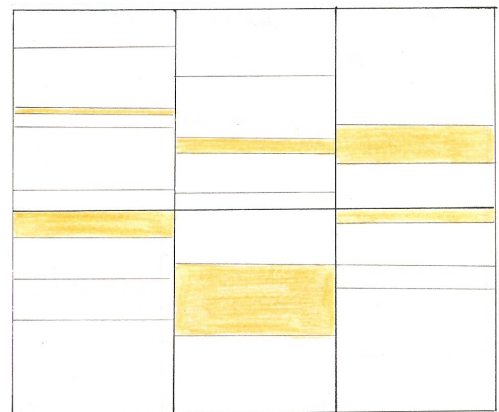
Each of these partitions has its own color and that also allows to visualize information. The partition representing the biggest bar has a color and the one representing the smallest has a different one. Those in between have colors in that are a mix of both to create a blending effect. This effect allows to understand information regarding contacts that stand-

out. For example, if from the smallest to somewhere in the middle there's a blending effect and then a big block of one color, is easy to understand that there is one person that the user contacted or contacted the user more.

When part of the area is selected, on a second area below is shown a zoomed in version of the selected area, now with a different division measure, using days. Still, the same principles apply. Also, on another area is shown a tag cloud with the themes used on the conversations on the area selected.

In figure 9 is possible to see part of what happens when we click a given partition, every partition that represents that contact gets highlighted. Also, concerning each highlighted partition, is provided information about the communication evolution, like how many more or less emails. Also, the tag cloud is updated to themes concerning only the chosen contact.

The figures don't show, but if a user wants to find a contact and highlight him or her, a textbox will be available to do a search by terms to try to find that information. That way, the user can sometimes even search for more than one contact and highlight both, being able to do some comparison analysis between the communication with them.



**Figure 9: Sketch of a selection in the proposed solution.**

### Possible Problems

One of the possible problems concerns the possibility of contacts with the same number of emails sent to and/or received from or also whose difference in between them is not visually relevant enough to see the difference and allow to click a partition. A possible solution is representing groups of contacts with an additional color and by clicking it, visualize the contacts in the group. Moreover, we don't see this as a drawback but more like a feature of our project: if the user is using the bars, he's looking for patterns not for a specific person or group of people, if the user is looking for people he should use the text in order to search for that person.

### 4. SUPPORTED TASKS

As stated in 1, our main objective is to mirror the interaction we have with others via email exploring who have we





Figure 8: Sketch of an overall view of the proposed solution.

been talking to as well as the "why" and the "how" for that answers.

Briefly, our prototype suggests the following tasks (presented in the form of questions):

- Who have I been talking the most/least for the last days/months/years?
- How does my relationship with **Joe** evolved over time? Why is that?
- Are my relationships to **Tom** and **Mary** correlated? How and Why?
- How do all me communications evolve over time? Why? When do I send/receive more/less emails?
- Do I communicate to a lot of people? Or mostly with a small restrict group?

Answers to these and other questions should be found by using our prototype.

## 5. TECHNOLOGY TO BE USED

Our solution is a web application prototype based on a web server, implemented by us, responsible for fetching and processing all the data. Besides that, the server will also generate all the necessary code for the graphical user interface. For testing purposes the server and client machine will be the same but, on the real world, the server could be deployed on some remote machine.

The server is implemented in python and the graphical user interface is built using HTML and javascript. The visualizations are based on the Protovis[1]<sup>3</sup> visualization toolkit.

## 6. REFERENCES

- [1] M. Bostock and J. Heer. Protovis: A graphical toolkit for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15:1121–1128, November 2009.
- [2] A. Perer and M. A. Smith. Contrasting portraits of email practices: visual approaches to reflection and analysis. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 389–395, New York, NY, USA, 2006. ACM.
- [3] A. Tat and S. Carpendale. CrystalChat: Visualizing Personal Chat History. In *HICSS '06: Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, page 58.3, Washington, DC, USA, 2006. IEEE Computer Society.
- [4] F. B. Viégas, D. Boyd, D. H. Nguyen, J. Potter, and J. Donath. Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. In *HICSS '04: Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 4*, page 40109.1, Washington, DC, USA, 2004. IEEE Computer Society.
- [5] F. B. Viégas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 979–988, New York, NY, USA, 2006. ACM.

<sup>3</sup>Available in: <http://vis.stanford.edu/protovis/>