Using an Ethnography of Email to Understand Distributed Scientific Collaborations

Sarika K Sharma, iSchool Syracuse University
Matt Willis, iSchool Syracuse University
Jaime Snyder, Information School, University of Washington
Carsten Østerlund, iSchool Syracuse University
Steve Sawyer, iSchool Syracuse University

Abstract
In this brief poster abstract we explore the finding from previous research that distributed teams collaborating on research use email to an overwhelming degree. This email is the source of collaboration and one of the central documents in the practice of doing science. We present an early idea of email focused ethnography and using visualizations to assist in the qualitative exploration of analyzing email communications. Of interest is the utility of different visualizations to inform follow up interviews of longitudinal fieldwork and data collection. Two such visualizations are presented and described. Along with the benefits of the techniques we describe some of the challenges.
Keywords: Ethnography, Email, Scientific collaboration

1   Introduction

Scientists assemble complex and idiosyncratic document infrastructures when collaborating in distributed teams. These document infrastructures are small in scale unlike traditional cyberinfrastructrue projects which are large in scale (Sharma et al., 2014). These small scale infrastructures use many different tools and consumer software packages in varying configurations in order to get scientific work done (Willis et al., 2014). Even though there is variety in the digital tools these collaborations use, during our investigation of how scientists use documents in their distributed research collaborations we were struck by the central role in which email played in all of the 22 different collaborations we observed (Kaziunas, Sawyer, and Østerlund, 2011). The role of email in collaborations of social scientists, information and computer scientists is to exchange documents, coordinate distributed meetings, and discuss project tasks and work. We found email to be so heavily embedded in these groups that a form of email based ethnography give unique insight into the work practices of the research collaborations we observed.

With IRB approval and the knowledge of participants we set up email filtering on the email accounts of select project collaborators. These emails were then copied to a secure email account for analysis by our research team. We applied methods of content analysis along with close reading of the emails and developed several categories to tag email interactions. These tags from each email were then visualized using the statistics package *R* using the *ggplot* library. These visualizations of email tags proved useful to show participants in follow up interviews and have been highly informative when analyzed in tandem with transcripts of initial interviews we conducted. The first collaboration we are tracking has over a years' worth of email collected. Email metadata is parsed and sorted into an excel file. Collected metadata include name of sender, name of receiver, subject, date, time, thread identification number, attachment name, and attachment number.

2   Ethnography of Email

Our reading and observation of exported emails followed traditional ethnography methodology by Spradley (1980). A graduate student from the research team conducted weekly observations of the emails paying close attention to the content of the body, attachments, and metadata of the emails. An extensive field note journal was also kept to adhere to best practices of writing memos, questions, and observations as the analyst became familiar with how a distributed scientific team coordinates using

email. Email as a site of ethnography provided us with an interesting look into the nuances of a collaboration and at the same time posed methodological challenges.

Weekly observations of email brought to fore questions related to the boundary of the project, the actors involved on the project, and the level of the researcher's understanding of the project. We found that that understanding project emails became an opaque task: the project boundary was blurred amongst the administrative, university, and personal email. In order to clarify our interpretation of emails, we conducted interviews with our participants to address some of the questions that arose over the course of few months. The coupling of email observations with interviews provided greater insight into the interpretation of our ethnographic data.

Tracking email provided us with a sense of how work was delegated amongst the team. We learned about the social organization of the collaboration, specifically we were able to spot early signs of governance and hierarchy playing out through email. It is evident from email dynamics that graduate students were delegated tasks that pertained to scientific activities such as literature review, summarizing information for papers in the form of citations and/or compiling data for needs of the principal investigator (PI). In contrast, two PI's on the team exclusively emailed each other involving grant writing, funding, and conference travel. It was evident through weekly email observations and interviews that management of teamwork was borne out of efficient management style of the lead researcher.

Understanding communication patterns of email also gave us insight to the hierarchy of the collaboration. For example, only PI's were included on certain scientific activities. Rarely, were students involved in the decision-making processes of the research project. Graduate students seemed to only interface with a principal investigator to coordinate weekly meetings and exchange work delegated to the students. It was also observed that one graduate student played a leading role which they used to manage other students.

3   Distributed Collaborative Project

In the interest of additional context, and to provide a key to help interpret the visualizations in figures 1 and 2, the following table illustrates the roles of the distributed collaboration we analyze. Projects emails were collected from the team shown in table 1, a distributed social science research team that consists of two research professors, five graduate students, and one technologist. Of the eight members 6 are female and 2 male. Three members of the group are geographically dispersed.

| Member Number	Organizational Role
Member 1	Professor/Co-Principal Investigator
Member 2	Director/Co-Principal Investigator
Member 3	Graduate Student
Member 4	Graduate Student
Member 5	Graduate Student
Member 6	Graduate Student
Member 7	Graduate Student
Member 8	Library Affiliate
Member 9	Technologist
Member 10	Professor

Table 1: Distributed scientific collaboration group member roles

4   Visualizing Email

In addition to content analysis another strategy we employ to understand qualitative email analysis is through visualizations. These emails were tagged based on inductively coded themes. Two major codes were prevalent in the emails. The first is scientific activities which consist of funding, grant writing, publications, literature review and data analysis. The second tag is the email genre which includes

concepts like meeting, coordination, discussion, and sharing work. Using a custom script for the *R* statistics package email metadata was structured and organized on a variety of plots and graphs. Meta data of use includes sender, receiver, attachment name, attachment number and tags for scientific activities and email genre.

Two figures have been provided (Figure 1 and Figure 2). Figure 1 displays the team member on the y-axis and the date on the x-axis. Each dot represents an email or an attachment. Team members are randomly positioned. This figure represents the bursts of email that are happening in the month of September.
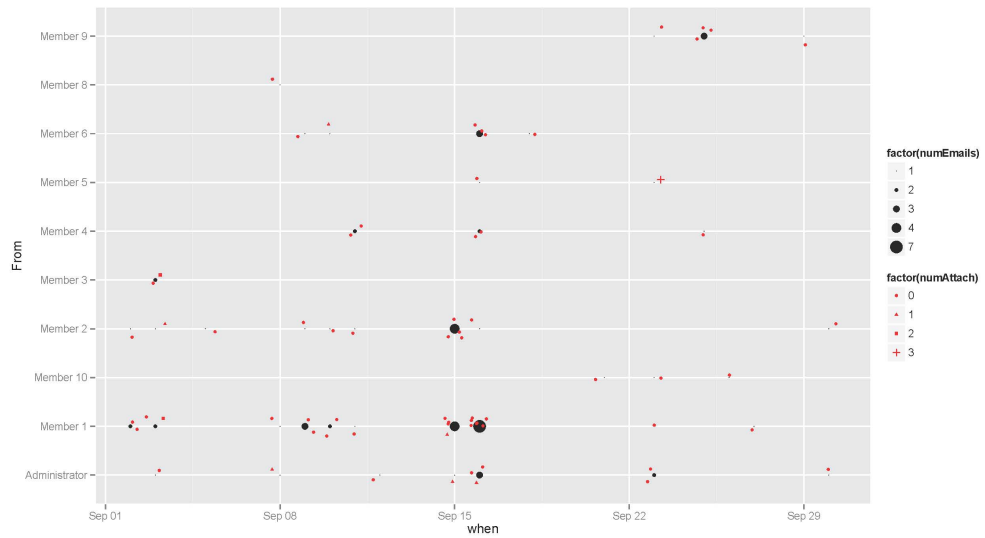
Figure 1: September 2014 email communication of a distributed scientific collaboration

Figure 2 displays team members on the y-axis and the scientific activity associated with the email on the x-axis. Each black dot represents an email and red dots an attachment. This visual displays the bursts of scientific activity that take place in the month of September.
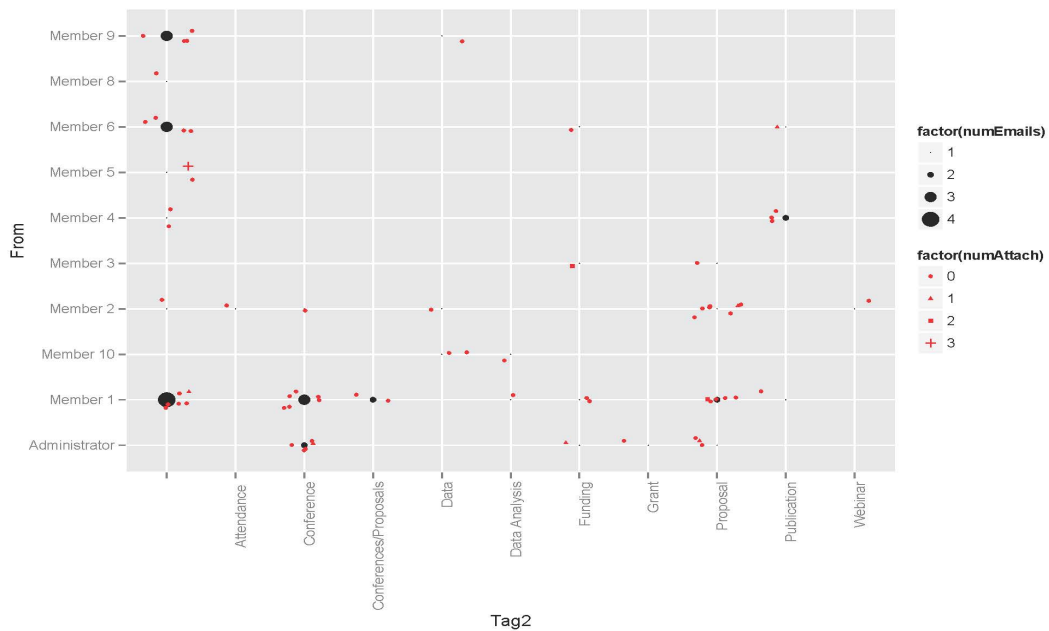
Figure 2: September 2014 email associated with scientific practices

3

We have found visualizations to be important as a pictorial representation of collaborative work that unfolds asynchronously over a long period of time through many messages. To facilitate follow up interviews with participants the visualization graphs will be shown as an elicitation tool to further interrogate in greater detail the collaborative work that takes place on email. For example, Figure 2 displays that team members 3-9 rarely send/receive emails about conferences and proposals. A follow-up question for our participants may be to ask about why it is that graduate students are not engaging in those email conversations.

5   Challenges and Conclusion

This technique of email ethnography and exploration of qualitative analysis through visualizations has informed our analysis and iterative data collection in pursuit of the project mission to understand the role of documents in distributed scientific collaborations. However, we have experienced equal challenges throughout this process. Chief among them has been gaining informants' agreement to allow the research team access to project emails. Even when access to project emails is obtained there exists a myriad of technical challenges in filtering incoming emails depending on the email client or email service used. Challenges also exist concerning the participants workflow. We use the built in inbox filtering and inbox rules that are available in the email client the participant is using. This is to avoid running cumbersome software or custom scripts which can interrupt the workflow of the participant. Our guideline for monitoring email has been to use it as an unobtrusive observation technique.

Similar to the discussion and method of trace ethnography using log data and other digital footprints generated by digital information technologies (Geiger & Ribes, 2011; Østerlund, Sawyer, Ribes et al., 2013), we want to explore the potential uses of email in research. Especially given the prevalence and extreme reliance of email in certain contexts, particularly in the doing of science. We have found the discussion on using email in research to be lacking. Specifically on including ways to incorporate email into the process of iterative data collection and analysis. We also think there is promise in the qualitative exploration of email using different visualizations.

References

Geiger, R. S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In *Proceedings of the 44th Annual Hawaii International Conference on Systems Sciences* (pp. 1–10).

Kaziunas, E., Sawyer, S., & Østerlund, C. (2011). Social scientists, documents and cyberinfrastructure: the cobbler's children or the missing masses?. In Proceedings of the 2011 iConference (pp. 166-173). ACM.

Østerlund, C., Sawyer, S., Ribes, D., Shankar, K., & Geiger, S. (2014). What to Do with All those Traces People Leave Behind: Computing, Culture, and (Bits of) Context? In iConference 2014 Proceedings (p. 1234 - 1237). doi:10.9776/14250

Sharma, S., Snyder, J., Østerlund, C., Willis, M., Sawyer, S., Brown, M., & Skolozar, D. (2014). Virtual Collaborations: Infrastructures That Support Document Practices. In *iConference*. Berlin, Germany.

Spradley, J. (1980). Participant Observation - Steps: 1, 2, & 3. In Participant Observation (pp. 38–72).

Willis, M., Sharma, S., Snyder, J., Brown, M., Østerlund, C., & Sawyer, S. (2014). Documents & Distributed Scientific Collaboration. In *CSCW*. Baltimore, MD.

Contents