UNIVERSITY OF GOTHENBURG

Designing a Visualization Application for Interaction Speed Improvement in Email Networks

Bachelor of Science Thesis in the Programme Software Engineering and Management

Eric Britsman
Fangzhou Cao

Designing a Visualization Application for Interaction Speed Improvement in Email Networks

Eric Britsman
Fangzhou Cao

Examiner: Lars Pareto

University of Gothenburg
Chalmers University of Technology
Department of Computer Science and Engineering
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Designing a Visualization Application for Interaction Speed Improvement in Email Networks

Fangzhou Cao, Eric Britsman
Software Engineering and Management
Dept. of Computer Science
Chalmers University and University of Gothenburg
Gothenburg, Sweden
{guscaofa, gusbriter}@student.gu.se

ABSTRACT

Hidden problems can be found in informal communication networks (such as email), and there is still much room for improvement on awareness of these problems. One area where such problems can be found is in the speed at which members of these networks communicate with each other (interaction speed). In this thesis, we have documented the Design Science Research study we have conducted in order to establish what causes these interaction speed problems in email networks, and how a visualization of the network could help solve them or reduce their frequency. Our research has resulted in us producing a feature list and a basic architecture for the visualization, as well as mockups based on those artifacts, along with scenarios that clarify what these mockups represent. We have also used our results to outline and discuss the main issue areas within organizational email networks. Finally, we provide guidelines for how our visualization application design can be used to reduce and/or solve issues within these areas.

The contributions of this study are important due to how organizational resource and communication routes (such as email networks) can be improved through visualization and analysis. We have also identified a knowledge gap regarding the design of email network visualization applications with the specific purpose of improving interaction speed, which implies that our results do not overlap with those of previous studies.

Keywords: Email network visualization. Interaction speed improvement, Social network analysis, Design science research.

1. INTRODUCTÌON

With the rapid growth of internet technologies, communication networks have been steadily transferred into internet-based communication such as email, instant messaging and various social network platforms. There have also been many papers describing the usage of these informal communication networks in organizational contexts (for example [10] and [14]). Knowledge of these networks and how they can be improved has great importance for successful management in an organization [14]. Meanwhile, hidden issues can be found in such networks, and a long-standing problem has been to deal with these issues and the awareness of their existence [10]. One area where such problems can be found is in how efficiently the members of these networks communicate with each other (interaction speed). Martini et al. [1] argue that there are several visible effects from the root factors for interaction speed slowdown that they have defined. Additionally, data has been presented in the literature [5], [17] which shows that a common procedure for analyzing informal communication networks is to visualize them. This then brings up an interesting question; can such networks be analyzed for interaction speed slowdown through the use of visualization?

Within the area of visualizing communication, the usage of email network visualization for different purposes has been especially well documented ([4], [10], [12], [14], [17]). The previous studies on email network visualization that we have found are mainly focused on studying the communication patterns ([12], [14]) and social relations ([4], [17]) of the people in the network. However, we have yet to find any studies related to the visualization of email networks that are focused on interaction speed improvement. This indicates that the results of such a study would not overlap with previous research. Since the email network at Gothenburg Computer Science & Engineering department (CSE) was perceived as slow by our main stakeholder, we also had an opportunity to study a real-world example of such a problem.

In order to resolve problems related to slow email communication, we have in this paper established how one could visualize an email network with the purpose of improving interaction speed. Part of this also included investigating the cause of slow interaction speed in general and how it has been dealt with in earlier literature. These goals of visualizing communication and understanding interaction speed are manifested in our two research questions:

- ***RQ1:*** *What issues cause interaction speed slowdown in email networks?*
- ***RQ2:*** *How can we design a visual solution for monitoring and improving interaction speed within email networks?*

Our research has led to several contributions:

- A feature list detailing the functionality we think the visualization would require based on analysis of our findings, as well as a basic architecture with suggested components.
- Several mockups and scenarios that explain and show how the visualization would work and look like (Appendix B: Scenarios & Mockups).
- Guidelines for how our designs can be used against the issue areas of email communication that we have identified (these issue areas are "Determining who to contact" and "Detecting presence of interaction speed slowdown factors and effects").
- By answering our research questions we have also contributed to filling the knowledge gap of interaction speed improvement through visualization of email networks.

This thesis is organized as follows: In the "Theoretical Framework" section, we give definitions and explanations of the major concepts related to our study (interaction speed & email network visualization), while also presenting previous research related to these areas in greater detail. In the next section (Process & Methods), we then provide information about every aspect on how we have conducted the study, for ease of replication. We point out the research site where this study took place, and then we describe what process was used to plan the study. This is then followed by an explanation of what methodologies were used to collect and analyze our findings. After that, we present our findings, filtered through the previously described analysis methods. We then discuss our results (and how they can be used), debate the validity of our study, and outline the possible opportunities for future work that we have identified. Finally, we have some conclusions on our project.

2. THEORETICAL FRAMEWORK

2.1. Interaction speed

Our view of email interaction speed (inspired by Martini et al. [1]) relates to how fast individuals (or other organizational units), respond to each other's' requests by email. Specifically, it is defined in our research as the linear negation of email response time: the time between the initial creation and sending of a request, and the response to said request. (This response time is calculated in work hours rather than actual hours, in order to account for weekends and holidays).

According to Martini et al. [1], interaction speed depends on several organizational, architectural, and individual factors that may or may not be managed. They have also identified ten such factors. These factors generate one or more interaction effects that can be observed in an organization (eight such effects were identified). They also provide seven recommendations for dealing with these factors, based around their area of communication between agile development teams. A few of these recommendations are also relatable to email communication in general. The factors, effects and recommendations that we have taken inspiration from are as follows (numbered as they were in the original source):

- **Factors: F1:** Knowledge unavailability, **F2:** Expert's reputation, **F3:** Unclear requirements, **F4:** Unexpected feature dependencies, **F5:** No co-location, **F6:** Lack of common time, **F7:** Mismatch of communication styles, **F8:** Slow resource indexing, **F9:** Low prioritized interaction
- **Effects**: **E1:** Waiting for communication, **E2:** Waiting for value, **E3:** Intense communication, **E4:** Corrupted communication, **E5:** High interaction frequency, **E6:** High task frequency, **E7:** Heavy interaction tasks, **E8:** Corrupted value.
- **Recommendations: R1:** Make experts available**, R6:** Shared Calendars, **R7:** .Creating awareness.

The exact definitions of each factor, effect and recommendation can be found in appendix A of this paper.
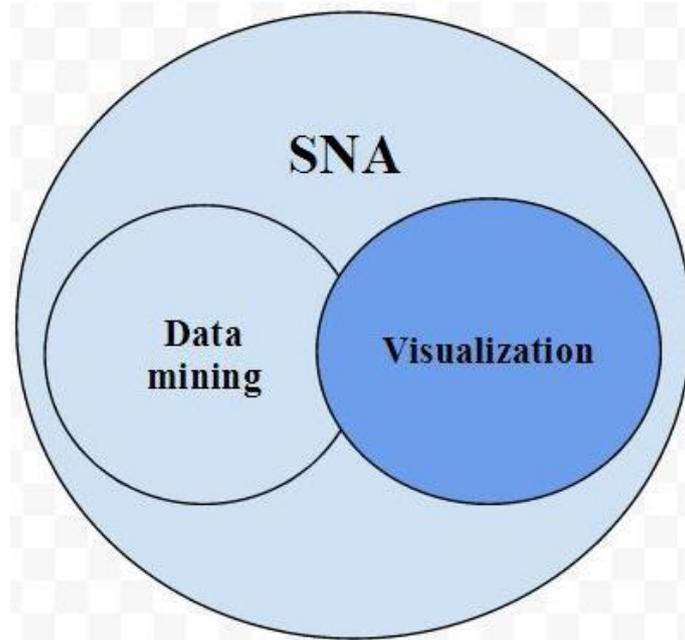
2.2. Visualization of email networks

Figure 2a. SNA overview diagram

Using social network analysis (SNA) is a common way of finding and resolving problems related to informal communication networks within organizations [17]. Email networks are an important branch of SNA that have been popular for both analysis and visualization, since the exchange of emails among individuals in organization is a good indicator of their relationships and responsibilities [14]. Communication networks are commonly visually represented using node graphs [8]. The nodes in these graphs represent the people and groups in the network, while the links show relationships or flows between those nodes [8]. In the case of an email network, each node represents an email address and each edge between two nodes represents an email exchange between those two addresses [17]. Analysis of these networks is applied in order to comprehend how members in email networks are connected and how those connections influence the network.

Our literature search for SNA visualization has brought to light that various methods for visualization of email networks for several purposes have already been researched and published. For example, in the research paper "Visualization and Analysis of Email Networks" by Fu et al. [17], they present four different methods: sphere drawing, hierarchical drawing, temporal display, and ambient display. Existing tools such as Thread Arc have been designed to help people use threads found in emails, which also combine the chronology of messages with the branching tree structure of a conversational thread in a mixed-model visualization [3]. Also, Fris et al. [10] have written about design principles that can be derived from the process of constructing and evaluating a real-time multi-user visualization tool for SNA.

Email network visualizations have also been used for different purposes. For example, Newman et al. [12] uses email network visualization to find patterns in the spreading of computer viruses, and Guimera et al. [14] uses email network visualization to observe informal communities in an organization. Another study by Christian et al. [4] uses email network visualization to compare social status and email activity to commit activity, while Fu

et al. [17] uses email network visualization to find communication patterns between different groups, to evaluate the evolution of changing relationships over time and to find social circles. Also, Fris et al. [10] uses email network visualization to help with organizing emergency response efforts. As previously mentioned however, none of these papers actually visualize communication with the purpose of improving interaction speed.

3. PROCESS & METHODS

3.1. Research site

As shown in **RQ2**, we wanted to research how to design an email visualization application for interaction speed improvement. We have also used a specific research setting to gather the data necessary for creating and evaluating such designs. This research site was at the Gothenburg CSE Software Engineering & Management program, since the problem identified there was seen as relevant to our research area, thus making it a good context to be studied. This site has affected who was interviewed, who is modeled in the visualization and who in the future will be asked to try out any prototypes (since most of these people are connected to the SE&M program).

3.2. Research process

Since our goals included the design of a visualization application, a research process that takes design into account was needed. This led to us researching different kinds of Design Research. By studying Action Design Research [11] we were able to conclude that the close, active collaboration implied by the "action" part wasn't necessarily applicable to our project. Instead, we chose to use Design Science Research, which is a research process based around using design and development as a tool for data collection and proving of concepts. As stated by Hevner et al. [2], Design Science Research is research that aims to create a purposeful IT artifact in order to address an important organizational problem. This definition fitted well with the purpose of this thesis, which is why design science research was chosen. The Design Science Research papers that were looked at for this study ([2], [6]) were also chosen due to how they both specify frameworks for conducting and documenting Design Science Research projects. Especially, Hevner et al:s paper [2] provides guidelines on the contributions that can be extracted from a Design Science Research project, which we have used as a reference for planning our own contributions. These contributions are:

1. **The designed artifact(s)**
   Most often, the main contribution of Design Science Research is the artifact that is used to try and solve the identified problem(s). In our case we planned to produce a feature list as well as mockups and scenarios in order to fulfill this category.
2. **The foundation of collected data filtered through analysis, and discussions/ conclusions based on it**
   In our case this category led to us realizing that we should also try to describe issue areas and guidelines for solving/reducing the risk of issues within these areas (as a way of discussing our results).
3. **The methodologies used for analysis & collection of data**
   This simply means that researchers should detail the procedures that were part of the project in question (for easier replication), which is also the reason for us describing our methodology extensively.

We also chose to use the Design Science Research Process (DSRP) (created by Peffers et al. [6]) in order to plan our research project (the details on how exactly DSRP was used can be seen step-by-step in Appendix D: DSRP).
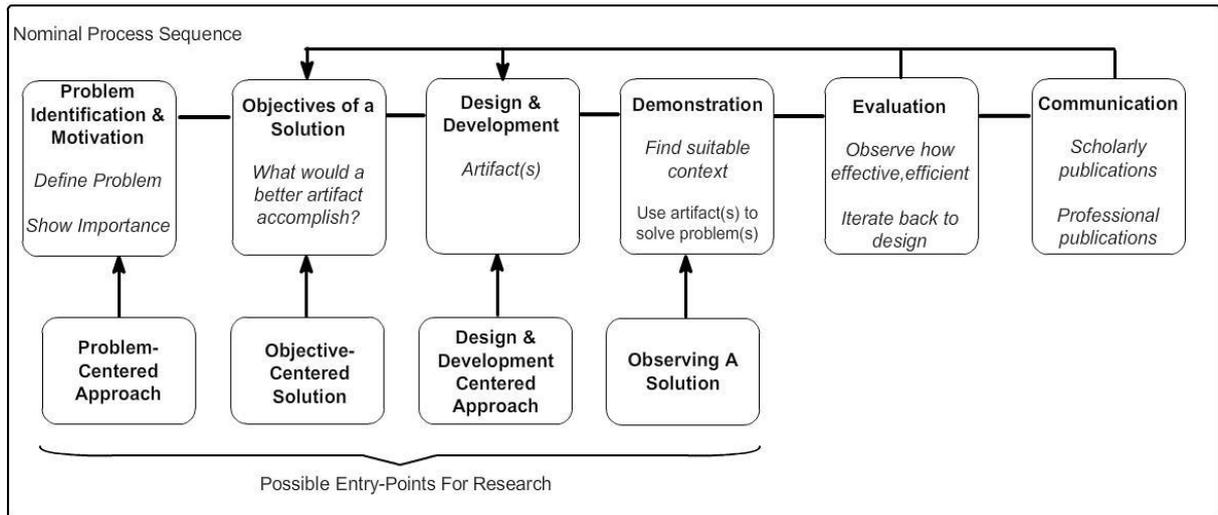
Figure 3a. DSRP

Part of planning the process was to divide it into three iterations. The focus of each iteration was based around what was deemed as necessary to fulfill before the main focus of the next iteration could be worked on. Five out of the six activities of DSRP [6] are represented within these iterations, and the sixth activity, communication, is represented by the production of this thesis, along with any presentations of the project. The planned iterations, and their main focus, were:

Iteration 1: Pre-study data gathering and initial feature specification creation
DSRP Activities:
1. Problem identification (literature review, semi-structured interviews)
2. Objectives of a solution (technology survey)

During iteration 1, our research problem and related knowledge domains were identified, and an initial list of feature ideas was also written using data gathered via literature, technology survey and the authors' own ideas. These features were then used as a basis for designing the first set of open-ended interview questions, which were used in three interviews. The results of the iteration 1 interviews were then also used to extend the feature list, and a subset of the features was identified as "core" for implementing in mockups/first prototype.

Iteration 2: Mockup design and evaluation
DSRP Activities:
3. Design and development (mockups & scenarios)
4. Demonstration (structured interviews)
5. Evaluation (structured interviews)

This iteration consisted of the design and creation of different mockups (with scenarios that explain them), which were then evaluated during interviews by representatives from the research site. An accompanying questionnaire was also produced, as support for the evaluation process. The mockup evaluations led to insight in areas where our current ideas and features were lacking, which led to changes and additions to the existing feature list and mockups. We then analyzed our final results in order to define issue areas and describe guidelines related to them.

Iteration 3: Prototype design and evaluation (was not reached)
DSRP Activities:
3. Design and development (proof of concept prototype)
4. Demonstration (structured interviews)
5. Evaluation (structured interviews)

This iteration is similar to iteration 2, except it is based around developing a fully functional prototype with all necessary components. During this iteration, various layout algorithms should be tested (as part of testing the API:s of the visualization tools we are interested in using), in order to decide what the optimal layout type for our visualization is. The features of the prototype will initially be displayed using static/simulation data, and it will be evaluated through structured interviews using a questionnaire derived from the one used in the previous iteration. This iteration was not reached in time for the deadline of this thesis. However, the design for the prototype is still visible through our contributions. (architecture and feature list in findings, guidelines in discussion, and the contents of Appendix B: Scenarios & Mockups).

3.3. Data collection methods

Literature review
In this study, literature review has been used to gain a better understanding of the problem area of interaction speed, as well as to help identify challenges and methods of visualizing communication. Online resources such as IEEE Xplore, ACM digital library and Google scholar have been used to search for literature. Selection was done on a per paper basis rather than by source preference.

Literature review search terms
The following keywords were used in our literature search:

- *Interaction speed*
  Measuring interaction speed, Measuring communication speed
- *Communication Analysis*
  Communication network analysis, Social network analysis, Email analysis, Email datamining
- *Visualization*
  Social network visualization, Communication visualization, Email visualization, Interaction speed visualization

Technology survey
Another part of this project was the gathering of data and identification of features by looking at existing technologies related to the research problem. Specifically, our technology survey has mainly been based around **RQ2**, that is to say how to visualize whatever information we deem necessary for interaction speed improvement of email networks. By studying technologies related to visualization of communication networks, tools/API:s have been found that are being considered for usage in future prototype development. Proprietary tools will be excluded from implementation in any future prototype, precisely due to their proprietary nature. These tools and their documentation have still been looked at however in order to establish which of their features we want to try and emulate. For mockup creation, we essentially looked for mockup tools that could be used to reasonably illustrate the feature ideas extracted during iteration 1.

Technology Survey search terms

The following keywords were used in our search for tools/API:s for visualization: Data visualization, Social network visualization, Communication visualization, Email visualization, Graph database visualization, Neo4j visualization, Force-directed layout, Hierarchical clustering, Centrality analysis, Online mockup creation, Wireframe mockup tool.

Interviews

Interviews were conducted differently between iteration 1 and iteration 2. The questions used can be found in Appendix E. As recommended by Hove & Anda [15], voice recordings were also made of these interviews, in order to ensure accuracy. The semi-structured interview was identified as a suitable interview type for the iteration 1 interviews, due to its reliance on open-ended questions (which according to Hove & Anda [15] is a great way to generate qualitative data, since it often leads to interesting follow-up questions). For iteration 2, while there was a lot of overlap from how we conducted our semi-structured interviews, we switched to using structured interviews for one specific purpose; the gathering of user feedback on the mockups we had created. The structure of the iteration 2 interviews consisted of a short presentation of our topic followed by showing and explaining our mockups and handing out a questionnaire.

Our choices on who to interview was based on advice from Myers & Newman [9] regarding how special care should be taken to interview "gatekeepers" (since the level at which the researcher enters the organization is crucial). We started by interviewing our main stakeholder (who is the program manager of the SE&M program at CSE). This stakeholder then provided several recommendations for interview candidates, based on his knowledge of the organization.

3.4. Data analysis methods

Literature analysis

The literature found through the literature review has been analyzed with the purpose of identifying ideas for features from those sources. From interaction speed literature we have extracted feature ideas specifying what kind of functionality the visualization could have in order to address and/or detect the factors, effects and recommendations from Martini et al. [1]. From visualization literature we have extracted feature ideas related to the appearance and data presentation of our visualization. The visualization designs found in these papers also helped influence design choices on the basic architecture of our future prototype.

Technology survey analysis

The technologies that have been surveyed were analyzed by comparing their functionality to the initial feature specification and architecture we had produced (determined via the literature analysis and our own previous knowledge and experience), in order to assess the suitability of these tools/API:s. They were also analyzed in order to identify new requirements based on their features for manipulating/interacting with diagrams, as well as on the types of diagrams they are able to create.

Interview analysis

Iteration 1:Semi-structured interview analysis

The semi-structured interview data was first analyzed by comparing the recorded responses to the questions that were asked. However, not all data from the semi-structured interviews in iteration 1 could be mapped to one of the initial questions. Instead, we grouped the responses into more general themes identified through the similarities we could find in the responses

from different interviewees. One such theme is "Strength of Email Communication", which refers to interviewee responses on what they like about using email. These responses were then used both to extract new feature ideas, as well as to provide support to previous ones.

Iteration 2: Structured interview analysis
Analysis of the data gathered from the iteration 2 interviews was done in a similar way to the semi-structured interviews. However, the themes used to group together responses from these interviews were based on the categories found in our questionnaire. These categories were based around the core concepts of our design, such as "interaction statistics" and "user information". We also added a category for general advice (for the few responses that we couldn't place in the questionnaire categories). The data from these interviews has been used to update and motivate our design choices, which are reflected in the final versions of our contributions.

Triangulated feature list
As emphasized by Runeson & Höst [13], triangulation is important to increase the precision of empirical research (especially when relying on qualitative data). Apart from achieving investigator triangulation by having both researchers participate in all steps of data collection and data analysis, we have also compared all the feature ideas that we were able to generate from our analysis process (in order to achieve data triangulation). Specifically, we analyze these feature ideas in order to create the finalized list of features for our visualization.

Scenarios & mockups
We finalized our mockups based on the previously mentioned feature list, and we wrote scenarios in order to analyze and explain how these features work together in certain situations (while also explaining what can be seen in our mockups).

4. FINDINGS
4.1 Findings from iteration 1 literature analysis
Interaction Speed Literature
We have summed up the feature ideas extracted from analysis of interaction speed literature in a table (*table 4a*), together with indicators on which of Martini et al:s [1] factors/effects/recommendations each idea is from. A more detailed account on the extraction of these feature ideas can be seen in Appendix D: Detailed feature extraction from factors, effects & recommendations.

Table 4a. Feature Ideas from Interaction Speed Literature

Feature Ideas	Factors, Effects & Recommendations
Connections between users should display interaction speed statistics.	F2, E1, E2
Role of each user should be specified.	F1, F2, F3, F4, F8, E6, R1
Manual input of data for users.	F1, F3
Ability to use filters to modify visualization appearance/content.	F1, F8, R1
Consider working hours and timezones for reply speed calculation.	F6
Have personal user pages with additional information.	R1
Show user-shared calendars	R6

Visualize unhandled emails.	F9, R7
Assign searchable keywords to users.	F4, R1
Visualize email priority.	F9, E1, E2, R7
Show information on user's current location.	F5
Ability for users to manually create social groups.	F7

Visualization Literature

A common trait of the visualization literature that we have analyzed in this study is that they all use the graph format for modeling communication (for example [10] and [17]). This led to us deciding that the email network should be represented in such a way, with nodes representing the people in the network, and edges representing the communication between them. This design decision is heavily reflected in our contributions, especially in the mockups we use to illustrate how the visualization application could look like. Some of these papers also led to more specific feature ideas based on how their visualizations work. These feature ideas are presented in a similar way to the ones from interaction speed literature, together with indicators on which paper the ideas originate from (*table 4b*).

Table 4b. Feature Ideas from Visualization Literature

Feature Ideas	Paper(s)
Log-files of email communication (email headers) can be used to construct email networks.	[14],[17]
Senders/receivers (their email addresses) that originate outside of the organization/company/group that you want to visualize should be filtered before the network is constructed.	[14],[17]
Multiple email addresses can be connected to the person they belong to via pattern matching algorithms.	[4]
Highlights benefits of making web-based applications.	[10],[16]
Size, colour, shape and location of graph elements can be used to enhance the visualization.	[10], [3]
The level of "importance" of a node can be emphasized using size.	[17]
Current "selection" in visualization can be highlighted via colour. Could also fade out elements unrelated to current selection.	[3]
The node graph could be movable.	[10]
Some situations may require manual input from users (which can then be reflected in the visualization).	[10]
Users could modify attributes of existing nodes/edges.	[16]
Nodes could represent other things than people.	[14],[16]
Input forms can be hidden to let the graph utilize the full display.	[10]
Detailed node/edge information can be hidden until they are interacted with.	[10], [3]
Available actions for users can be linked to clickable icons.	[10]
"Free-text" fields can be used for input of additional data.	[10]

Users could have the ability to zoom in/out on the visualization.	[10]
Timespans could be used to affect what data is visualized.	[10]
Variation on nodes can be used to represent people from different groups.	[14]
The visualization could be filtered by searching through people (nodes) based on certain attributes.	[7]

4.2. Findings from iteration 1 technology survey

Two major architecture/design choices affecting future prototypes (and the look of our mockups) were determined early during iteration 1 of this project:

- **Graph-based visualization**

  This decision happened early on during the literature review, due to how most visualizations of communication encountered were represented as node diagrams. When we started researching visualization tools it was also quickly discovered that these tools in general expect data from graph databases in order to generate diagrams. Thus, it was also decided that any future prototypes should use a graph database rather than a relational one. Based on evaluations on Java-compatible graph databases found in [16], we are currently interested in using Neo4j for that part of our architecture. The main reason for choosing Neo4j was the level of documentation compared to other graph databases, as mentioned by Huang & Anton [16]. Our own research into graph databases indicates that this documentation advantage is still relevant today.

- **Web-based visualization**

  In order to maximize accessibility and remove any needs for distribution or user-side setup, the visualization should be web-based. A web application also has the added benefit of having a highly customizable UI. More specifically, for this initial prototype we are planning to use the Maven web application framework together with HTML5, due to how one of the researchers is already familiar with it, along with its ease of learning/use. Maven would be used with Java, again due to the researchers' familiarity.

After the components mentioned above had been decided on, the rest of the technology survey mainly dealt with finding technologies to help visualize a graph database (rather than draw everything from scratch), with a particular focus on findings tools that work with web-applications**.** We found that the identified visualization tools that are based on JavaScript (D3, Infovis, Sigma) fit nicely with our previous choices. These JavaScript based tools also provide many interesting features for customization, manipulation and animation. We have also created a table that summarizes this part of our technology survey (*table 4c*).

*Table 4c.**Visualization technology survey summary*

Tool	Description	Analysis
JavaScript D3.js	A JavaScript library for manipulating documents based on data. It uses web standard such as HTML, CSS and SVG to visualize data.	Interesting features such as APIs related to Forced-Directed Graph, drag and drop, animated nodes and edges. Has an extensive gallery of examples for visualization customization, with some of these examples being interactive.

Neo4j Server Web Interface	The Data Browser Tab offers a handy visualization of your graph data. Users can select the nodes to be shown by id, index lookup or cypher query. A style editor will adapt the visualization to users' needs.	Extended capabilities of customizing the visualization. Uses Cypher query for searching nodes. Useful when starting with the backend part of the prototype. Just by creating some edges/nodes they can instantly be looked at with this.
Linkurious	A proprietary web-based application for searching and visualizing graph databases.	Has very interesting search engine for finding nodes and connections. Has a feature named "Stay focus" - a feature that helps to stay in control by focusing on the data related to users' search terms.
Keylines	A proprietary JavaScript toolkit for visualizing networks. It works in all major browsers, and on smartphones/tablets. It uses HTML5.	Has many animation features. Supports using custom images/stylings for nodes/edges.
Gephi	An open-source, platform independent desktop application for interactive visualization and exploration of all kinds of networks and complex systems, dynamic and hierarchical graphs.	Apparently has great performance, supporting networks up to 50,000 nodes and 1,000,000 edges. Has support for hierarchical graphs and clustering of nodes based on custom attribute.
Graphviz	An open source graph visualization software.	Mostly interesting for suggestions on how to style nodes/edges based on examples in the gallery of the tool's webpage.
ZGRViewer	A graph visualizer implemented in Java, specifically aimed at displaying graphs expressed using the DOT language.	Not fit for use with our purpose due to lack of features, but has an interesting "magnifying glass" feature.
Infovis	JavaScript toolkit which provides tools for creating Interactive data visualizations for the web.	API is not as large as D3's, and documentation level is also lower. However it shows many interesting animation features and diagram types in its interactive demos.
Sigma.js	An open-source lightweight JavaScript library to draw graphs, using the HTML canvas element.	Lacking in documentation but showed some interesting features in its interactive examples, such as hovering over nodes to reveal additional information.

4.3. Findings from iteration 1 semi-structured interviews

When listening to our interviewee responses, we identified three general themes that these responses could be grouped in. The first of these themes (**Descriptions of Email Communication**) groups together all responses that describe the email network and email communication process at CSE Gothenburg, based on interviewees' perception of and participation in said email network. The second theme (**Strengths of Email Communication**) groups together the responses that indicate the advantages of email compared to other forms of communication. The last theme (**Interaction speed of Email Communication**) focuses on interaction speed "problems" at CSE, and possible solutions that were suggested by the interviewees. The results of this first round of interviews have been used both for verifying the importance of email communication at CSE, as well as for drawing some parallels between interviewee answers and certain factors and effects found in [1]. The interview responses are shown below (*table 4d*), grouped into these identified themes.

Table 4d. *Iteration 1 Interviews*

Theme	Description	Responses
Interaction speed of Email Communication	Responses that describe the interaction speed of email networks in CSE department and possible solutions to promote it. For example what new features could be added to the existing email system to achieve better performance.	1. Email networks should be search-based, for example, search emails by name. 2. The role of each person in the network should be specified. 3. Access to a specific group of people, for example, access to the students who are applying for thesis approval in 2013. 4. Detailed subject and full identity of sender. 5. Use rules to filter email into manageable subsets. 6. It's good to know the email frequency, efficiency of colleagues.
Strengths of Email Communication	Strengths or beneficial effects of using email in CSE department compared to other communications such as phone calls and physical meetings etc.	1. Emails sent to the administration email address are handled among different administrators. 2. Email makes it easy to reach people for communication when working with two different campuses. 3. Email provides searchable logs of previous conversations.
Descriptions of Email Communication	Descriptive responses of email communications in CSE department, based on interviewees' daily perception and participation in email network.	1. Email is primary communication tool with both students and colleagues. Sometimes phone calls are used instead. 2. Mailing lists are used to reach large quantities of students at the same time. 3. Emails are used to get info from teachers overseas and get reply usually at night. 4. A central office stores all emails sent back and forth.

4.4. Findings from iteration 2 structured interviews

The second round of interviews was mainly used to get acceptance data on our mockups from the interviewees, as well as to improve/extend our existing features based on their responses. One way that these interviews affected our design is how they made us realize that our initial feature list did not properly take security and privacy into account. Thanks to one of our interviewees in particular we also gained a better understanding of what we could legally show in the visualization, which in turn led to the simplification of some features, and changes that make various pieces of information hidden unless a person manually gives consent to show it.

Qualitative analysis of the data gathered from the iteration 2 interviews was done in a similar way to the semi-structured interviews, and it is also displayed in a similar table (*table 4e*). The themes in this table are however based on the categories found in our questionnaire that was used during those interviews. This table also indicates changes and new additions made to our design based on interviewee responses (the number on each change corresponds to the number of the response that caused it).

Theme	Description	Responses	Changes Made
Basic Visualization	This mockup focused on showing our basic design for the email network graph.	1. Using dash/dotted line for representing difference in edges. 2. Look at Sim city traffic graph visualization for inspiration. 3. Visualization should be color-blind accessible. 4. The color choices for representing edges could be changeable.' 5. Colour for speed on edges could be based on colour intensity within one "group" of colours. 6. Consider which features are essential and which should possible to enable/disable by logged in user. 7. Use different icons for employees and students.	3. & 5. We changed the color of edges to blue color family and nodes to black due to color-blind issue. 4. User is able to customize colors for representing edges. 7. Icons were changed for employees; students still use the same one as before.
Interaction Speed Statistics	This mockup focused on showing what information we thought was relevant for measuring interaction speed between people, and how/where we wanted to show it.	1. Possibly use tag clouds to represent node tags. 2. Reply volume should be represented. 3. Statistics should be more intuitive. 4. Show more data on nodes rather than edge (easier to select). 5. Have the system indicate possible slowdown areas (and what factor/effect could be causing it) in the actual visualization. 6. Reply speed measurements should be based on work hours.	2. We represented each person's contribution in single email conversation. 3. We modified the statistics from just raw fact numbers to words like "fast", "medium" or "slow" etc. 4. Recent conversations were moved from viewing on edge to viewing on node.
User Page	This mockup focused on showing what information we thought was useful to know about each person in the visualization, and how/where we wanted to show it.	1. Have privacy settings to determine who can view your events on calendar etc. 2. Opening hour (for study administration etc) should be available. 3. Can shared calendars and location information lead to privacy issues? 4. Using pictures for people may be a bad idea, may cause bias when choosing who to contact.	2. It can be solved by having personal text/statements on user page.
Tag System	This mockup focused on showing how we thought tags could be used to easily identify a person's interests and area of expertise.	1. System should give suggestion on tag keywords when inputting. 2. Having categories for tags & have more specific tags. 3. Too much work to manually add tags to recent conversations.	1 & 2. Added feature for default tag lists for choosing existing tag instead of making new. Tag list should have both general and specific tags to choose from. 3. We added one feature that auto-generates predetermined tags based on user's keywords (for recent conversations).

Email Conversation Information	This mockup focused on showing how we want to visualize recent conversations, their priority/status, and show interaction speed statistics on a single conversation basis.	1. Email conversations visualization could be connected to outlook. 2. Single email conversation should include information on if the email is replied or not by the selected node. 3. Single email conversation should have different levels of priority on students and employees.	1. The information visualized for email conversations is extracted from email header files. 2. We represented the email status as "Handled" if at least one email is replied by the selected node. 3. For a single email, there are four different levels of priority: "urgent", "follow-up", "normal" and "don't reply". Employees can use all four levels. "Urgent" is not available to students.
Filtering Features	This mockup focused on showing off the different filters that we plan on implementing, as well as how the filtering effect should actually look.	1. Timespan filter could be handled via a dynamic slider.	1. The filter "Time span" could be controlled by time slider instead of calendar selector. (if performance allows)
General Advice	We also received some general advice that wasn't strictly related to the mockup categories/ questionnaire.	1. Look into integration with functions from currently used email clients. 2. Legal issues must be considered, especially in regards to information on students. Features must comply with these rules. 3. People who are on vacation could somehow have this status affect the look of their node, so people see they might need to find someone else to contact. 4. Does the system have to broad a focus? Does it include too much things outside of interaction speed improvement? 5. Possibly separate feature areas into separate views in the visualization.	2. Information shown on people in the visualization (via their node/user page) will be restricted if that person is a student, until they have manually given consent for it to be shown via managing their profile. 5. Userpages (from where you can access account management as well) will open in a separate page/tab from the actual visualization.

4.5. Triangulated feature list

In *table 4f* we show the full list of features generated through analyzing the feature ideas found during this project. We present both the feature name and a short description for each feature. We also illustrate which parts of our analysis these features are supported by, as well as how we have prioritized each feature for an eventual prototype implementation. Each feature can be traced back to its sources by looking for certain abbreviations, which are:

- Interaction speed literature analysis (**ILA**)
- Visualization literature analysis (**VLA**)
- Technology survey (**TS**)
- Iteration 1 interviews (based on email communication in general) (**I1**)
- Iteration 2 interviews (based on mockup feedback) (**I2**).
- Features from realizations we had while creating/updating our mockups (**MC**).

The three priority levels displayed in the table are as follows: rank 1 indicates both core features necessary for visualization of interaction speed, as well as the visualization features that we are the most interested in implementing. Rank 2 indicates some more advanced visualization features that wouldn't be implemented in the first prototype due to difficulty or less relevancy. Rank 3 is for highly optional features that probably won't be implemented. These prioritizations are separate from how our mockups were created, since implementation difficulty does not necessarily translate to being hard to show in a mockup and vice versa.

Table 4f. *Triangulated Feature List*

Feature(priority1>2>3)	Description	Source
FT1: Construct Email Network(1)	The ability to construct and store graph representations of email networks by reading log-files of email headers.	**VLA,TS**
FT2: Web-based Solution(2)	The ability to display email network visualization in a browser.	**VLA,TS,I2**
FT3: Use of Existing Authentication Systems(2)	The ability to log in to the webpage where graph is visualized through integration of existing authentication systems. Existing authentication systems also help to restrict access to only relevant people.	**I2**
FT4: Show Communication in Both Directions(1)	Visualization shall display all communication links between people. Two links per connection is needed (since the interaction statistics vary for each direction).	**I1,I2**
FT5: Using Colours & Shapes to represent information(1)	Colours/shapes/size of nodes/edges should be used to represent levels of certain attributes, for example: • Representing the interaction speed level of a connection (edge) via colour. • Differentiating people (nodes) in the graph by size based on how much they communicate via email.	**VLA,TS,I2**
FT6: Coulourblind-adjusted Colour Scheme(1)	The default colour choices in the visualization should be well suited for any possible colour blind users.	**I2**
FT7: Icon variation for major roles(1)	Students and employees should have different node icons to make it easy to differentiate the two at a glance.	**VLA,I2**
FT8: Combining nodes by email address(2)	Users should be able to connect their multiple email addresses to themselves manually. This will combine edges that lead to the same node (and the nodes these edges originate from), and reevaluate the interaction speed statistics between these people.	**VLA,I2**
FT9: Zooming(1)	The ability to zoom in/out on the visualization.	**VLA,TS,I1,I2**
FT10: Position-Indicating Mini-Map(2)	There should be a position indicator UI component that can be used to estimate in which quadrant of the visualization you are currently looking.	**I2**
FT11: Icon Interface(1)	The available actions for users should be linked to clickable icons. This includes settings, profile management, in/out zoom and graph help/explanation.	**VLA,TS,I1,I2**
FT12: "Hidden" Detailed Information(1)	Nodes/edges should show more information when interacted with (hovering/clicking).	**VLA,TS,I2**
FT13: Information Panel(1)	Additional information such as mentioned in **FT12** should be shown in an overlaying information panel at the right side of the screen, in order to avoid the obscuring of other elements that happens when the panel is placed on/near the element itself. This also means that the visualization can be scrolled in any direction without losing view of the information panel. If no element is hovered over, the box shows the data on your latest selected (clicked) element (if any).	**MC**
FT14: Selection Emphasis(1)	The user's current selection should be enlarged in order to make it stand out.	**TS,I2**
FT15: Quick-select Self(2)	The user should have a shortcut for making their node the currently selected element at any point.	**MC**
FT16: People Panel(1)	Below the information panel there will be a people panel. This people panel shows all the names of the people within the current filter settings. Clicking one of the names in the people panel makes the view center in on that node.	**MC**
FT17: Hideable Components(2)	Overlaying components should be individually hideable to let graph utilize the full display.	**VLA,TS,I2**

FT18: Colour Customization(3)	Users should be able to change the colours used for nodes, filters, highlights, reply speed levels and page background in order to suit their preferences.	**VLA,I2**
FT19: Overtime Interaction Statistics on Connection(1)	Edges between users should specify different overtime interaction statistic totals when selected.	**ILA,I1,I2**
FT20: Interaction Statistics on Recent Conversations(2)	Whenever a node is selected the information panel should also gain a button that can be used to switch between showing contact information and recent conversation data. Conversations with emails that reference message id:s outside the current timespan will be excluded from statistics calculation and from the users' recent conversations lists. This feature can only be accessed on a per user basis, by selecting a user and switching to conversation view in the information panel. Information on date, time, contribution percentage, status and reply speed will be available for each user's five most recent conversations (actual topic/contents excluded for privacy).	**ILA,I1,I2**
FT21: Tags for Recent Conversations(2)	Conversations will have automatically generated tags based on keywords found in email header.	**I2**
FT22: Recent Conversation Priority(2)	There should be an automatic system for assigning and visualizing priority levels on these recent conversations (also based on keywords in the email headers.) The email priority system should contain several priority levels such as urgent-important-followup-normal-low priority.	**ILA,I2**
FT23: Conversation Priority Restrictions(3)	Some priority levels should be restricted to certain people (for example highest priority should only be generated on conversations between employees). This can be verified by the system looking at the roles of the people connected to email addresses used in the conversation.	**I2**
FT24: Recent Conversation Status Icons(2)	These recent conversations will also have icons related to their status displayed beside them. This icon is an open envelope if you have contributed at least one message in the conversation. Otherwise a closed envelope icon will be used. Interacting with the mail icon on recent conversations should show in text the status and level of priority for that conversation.	**ILA,I2**
FT25: Additional Information on Node Selection(1)	The name, role and email address of each person represented in the network should be viewable when their node is selected. Email addresses will most likely be hidden on student nodes until they give consent for it to show (due to legal reasons). This should not affect employee nodes however, since their data is already available on other public sites. This additional information should also include the user's total average reply speed to each specified node category (for example employees and students).	**ILA, VLA,TS,I2**
FT26: Node Tag System(1)	It should be possible to manually add searchable tags to nodes. The node tag system should include both list of selectable predetermined tags, and the ability to create new tags (that need to be approved by an admin before they are added). Approved tags are also added to the existing list. The node tag system's list of default tags should contain a mix of general and more specific tags. For example, a user could choose to use either the "course registration" tag, or the more specific "course registration:GU" tag.	**ILA,I2**
FT27: Node Tagging Restrictions(2)	Users should be able to tag other's nodes (and will again need to be approved by admin if tag is new), but these tags must also then be approved by the user being tagged before they are added.	**I2**
FT28: Custom Groups(2)	Users should be able to create and manage their own customized group and then invite other people to it. Custom group names must be accepted by admin before the group is created.	**ILA,I2**
FT29: User Page(1)	There should be user pages for each person with both the information mentioned in previous features, and some additional information. These pages are reached by clicking a node and then clicking the name in the information panel. These user pages should open up in new tabs separate from the email network graph page.	**ILA,I2**
FT30: Free-text for adding information(1)	The user page should have a "free-text"-field allowing the user to manually add additional information about themselves.	**ILA, VLA,I2**
FT31: Location Information(1)	Information regarding a person's location (for example office or classroom) should be available on the user page. This feature could possibly be linked to existing maps. This feature is for when people want to contact the other person directly.	**ILA,I2**

FT32: Shared Calendars(1)	There should be optional calendars on the user page, preferably connected to existing ones such as Google calendars. The point here is to "share" the calendar with the visualization instead of on a per person basis.	ILA, VLA,TS,I2
FT33: User Page Management(1)	Users should be able to manage their pages (accessible through same tab as user page tab). This includes linking calendars, adding/updating/removing information etc.	I2
FT34: Filtering Visualization(1)	There should be various filters for altering the visualization. These filters should be usable both separately and together. The filters we have currently identified as suitable are: • Node tag filter. • Recent conversation tag filter. • Node relations filter (to only show people you have communicated with). • Custom group filter (dropdown list with content based on the group membership of the currently selected node). If no node is selected then all groups are selectable. • Timespan filter based on choosing a start and end date.	ILA, VLA,TS,I1 ,I2
FT35: Highlighting Recent Conversation(2)	Clicking on a recent conversation highlights relevant nodes and edges. Also alters the people panel to only show the people affected by this highlighting and adds their statistics for this conversation next to their names.	MC
FT36: Partial Filter/Highlight(1)	Highlighting/filtering system changes/removes the colour of the irrelevant (filtered/unhighlighted) nodes and edges (rather than hiding those elements completely). This in turn makes the relevant elements appear highlighted since they remain unchanged.	I2

4.6. Scenario & mockup example

As previously mentioned, we have created several mockups in order to visualize our features, followed by writing scenarios that describe what exactly can be seen. Below, we show one of these scenarios (and its accompanying mockups) as an example. The rest of the scenarios and mockups (and information on how they were created) can be found in Appendix B.

Scenario: Searching with keywords
Description:
This scenario shows how a user can attempt to find relevant nodes through filtering the visualization by searching for certain keywords. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT17: Hideable Components**
- **FT34: Filtering Visualization**
- **FT36: Partial Filter/Highlight**

Priority:
High
Preconditions
User is already logged in.
Postconditions
The user's current view has been altered in order to highlight elements relevant to the tag(s) he/she has searched for.

Actors:
User

Basic flow:
1. User inputs keywords in search bar.
2. User selects node tag filter.
3. User starts searching.
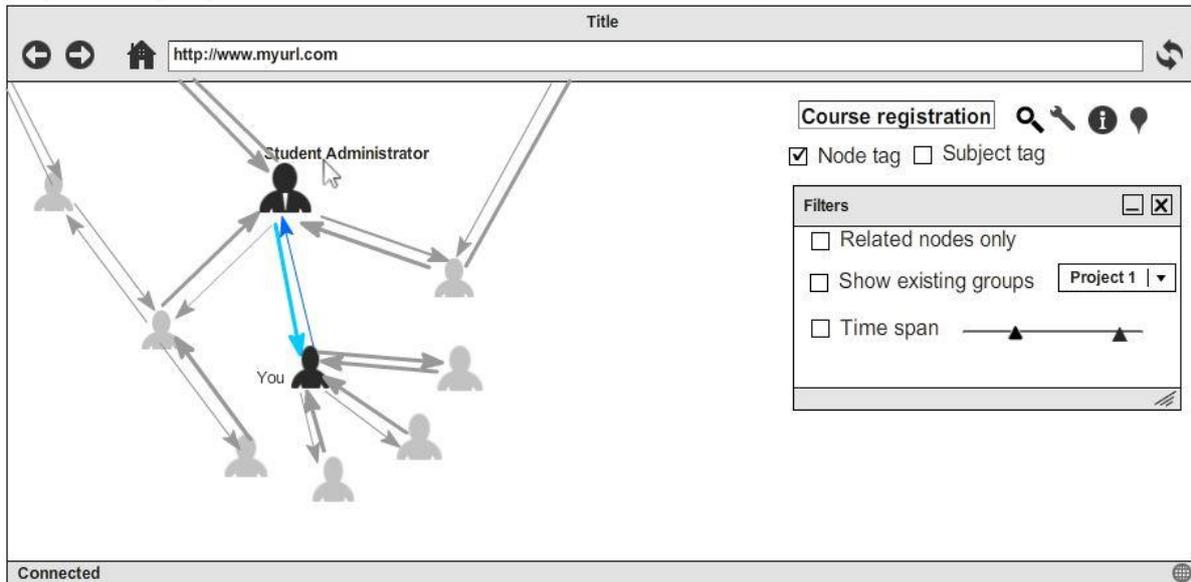4. System highlights the search result in the visualization (*4f*).

Figure 4g. Node tag search mockup

Alternative flow 1:
2a. User selects subject tag filter.
- 2a1. User starts searching.
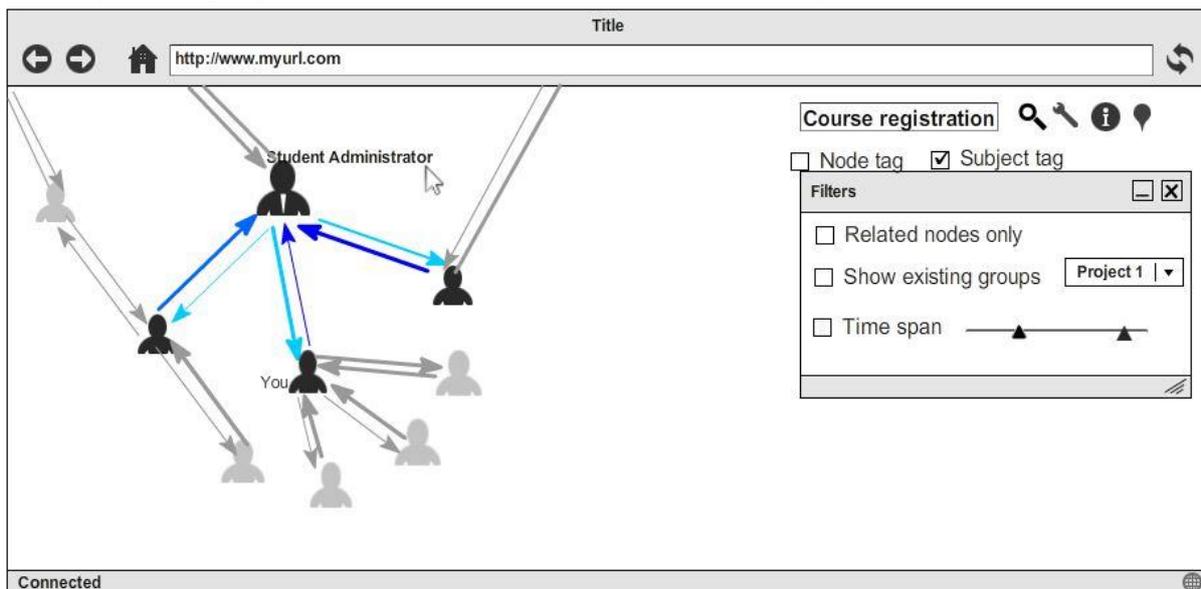- 2a2. System highlights the search result in the visualization (*4g*).

Figure 4h. Subject tag search mockup

Alternative flow 2:
4a. System returns "no result found."

5. DISCUSSION
5.1 Issue areas & guidelines
In order to answer **RQ1:** *What issues cause interaction speed slowdown in email networks?*, we have identified issue areas within email communication (based on the issues we have identified during our data collection and analysis). With issue areas, we refer to collections of similar and/or related issues grouped together into a general theme. We also use examples based on features generated from our findings in order to provide guidelines on how to solve and/or reduce the risk of issues related to these areas (as part of **RQ2:** *How can we design a visual solution for monitoring and improving interaction speed within email networks?*). These guidelines in turn helps motivate our design choices (by showing the purpose of our designs). The features not found in the guidelines of this section are features centered on graphical design suggestions for future visualization prototypes (such as **FT5: Using colours & Shapes to represent Information**), or aimed at certain quality attributes (such as **FT2: Web-based Solution** for accessibility, **FT3: Use of Existing Authentication Systems** for security and **FT6: Colourblind-adjusted Colour scheme** for usability). These types of features are not strictly related to interaction speed improvement or "issue-solving".

By focusing on finding these issue areas and giving general guidelines, we have provided contributions that are relevant to improving interaction speed in organizational email networks in general (rather than just for improving the email interaction speed at Gothenburg CSE). Based on the issue areas we identified, we also believe that they can be applied to visualizations of other types of informal communication networks (rather than email only). We say this because we believe that these issue areas are general enough to not be platform specific.

Issue area 1: Determining who to contact
By going through our interview findings, we noticed that most of our interviewees from iteration 1 mentioned issues with emails being sent to the wrong person. They also commented on the slowdown this creates by forcing you to forward these emails to a third party (and determining who this third party should be). These issues are related to accessing the correct human resource needed for processing those email interactions. Tracing back to the findings from factors, effects and recommendations [1], some of the factors (**F1**-**F4**, **F6** & **F8**) can be connected directly to these issues. By analyzing these connections, we drew the conclusion that accessing the right person in an organization is a major issue that affects email interaction speed. Thus, increasing the chance of accessing the right person in an organization should also improve email interaction in that organization's network. In the coming section, we provide guidelines for how some of our features can be applied to this issue area.

Guidelines for determining who to contact
Users in the organization being visualized can utilize **FT25: Additional Information on Node Selection** to gain knowledge of the roles of each individual, which should make it possible to increase the chances of finding the right person to email. Also, If users add their own information on the free text area (**FT30: Free-text for adding information**) available on their user page (**FT29: User Page**), this information can then be viewed by other users so they can try to decide if this is the person that he/she actually needs to contact. The information from these features help against *F1:* Knowledge unavailability*, F3:* Unclear requirements*, F4:* Unexpected feature dependencies and *F8:* Slow resourcing indexing, and it can also help offload interaction from a single person as in *F2:* Expert's reputation, since it should be easier to locate someone else with similar expertise. These features are particularly useful after the amount of nodes in the view has been reduced through usage of one or several of the filters

mentioned in **FT34: Filtering visualization**, or if the people in a certain conversation have been highlighted using **FT35: Highlighting Recent Conversation**.

Furthermore, users can add their available time slots to the **FT32: Shared Calendars** on their personal page which helps against *F6:* Lack of common time. By viewing these calendars, people should be able to set realistic expectations on the available time for daily meetings and other working related issues. Seeing calendars shared with users should also indicate when a certain user most likely won't be answering emails (since they have something on their schedule for that time). If this information is available, instead of waiting for response from one request handler, the email initiator could potentially send out the request by email in parallel to different people that are involved in processing the request, instead of relying on others to involve the necessary people. We believe that **FT34** also can be used to narrow down the people that would possibly be involved in processing the request, making this task easier.

In addition, users can use **FT26: Node Tag System** to add tags to themselves and others. By applying the node tag filter from **FT34**, other users in the organization can then manually input these keywords, in order to target people with certain tags. Similarly, users can search keyword(s) with the conversation tag filter from **FT34**, based on tags generated by **FT21: Tags for Recent Conversations**. This would then let a user see a highlight of all people that recently had conversations which generated that keyword as a tag. Users can also use **FT28: Custom groups** to create custom groups that both themselves and others can filter the visualization with, by using the custom group filter mentioned in **FT34**.

Issue area 2: Detecting presence of interaction speed slowdown factors and effects
By going through the factors and effects from Martini et al. [1], we also found direct connections to how showing information on user interaction could be used to imply the presence of several of the issues outlined in that paper (**F2, F6, F7, F9, E1-E3, E5 & E7**). This information includes reply speed, email quantity and recent email activities. To help with this area, we created features that specify what interaction statistics need to be visualized, and how to visualize them, so that the visualization can be used to identify where in the email network these factors/effects may be present. We also used the provided recommendations **R6:** Shared calendars and **R7:** Creating awareness as feature inspirations related to this issue area. In the coming section, we provide guidelines for how those features can be applied to this issue area.

Guidelines for detecting presence of interaction speed slowdown
In general, being able to accurately see where communication is slow can indicate **F6:** Lack of common time*,* **F7:** Mismatch of team's styles of communication*,* **F9:** Low Prioritized interaction*,* **E1:** Waiting for communication and/or **E2:** Waiting for value*,* and if it's frequent it could indicate **F2:** Expert's reputation*,* **E3:** Intense communication*,* **E5:** High interaction frequency and/or **E7:** Heavy interaction tasks. If the user already has a circumstantial connection in the visualization to the person they want to email, they can use the communication link from **FT4: Showing communication in both directions** that goes from the email receiver to them in order to estimate the waiting period for that person's response. This is done by using **FT19: Overtime Interaction Statistics on Connection** to look at statistics such as reply speed for that communication link. If the two users have had a conversation recently, the first user can base their estimation on information from **FT20: Interaction Statistics on Recent Conversations** instead. If the link between these two users is "weak" (low email quantity) or nonexistent, the first user can instead use **FT25: Additional**

Information on Node Selection to look at the other user's node for more information, and use this user's average reply speed to the first user's node category in order to estimate when a response will be received. The first user could also look for **FT32: Shared Calendars** (based on **R6:** Shared calendars) on the other user's user page, and use that calendar to see when the other user most likely won't be able to answer any emails. These features should show when there is a risk that the response will take a long time, which then can lead to the first user contacting someone else in order to avoid **E1:** Waiting for communication and **E2:** Waiting for value.

As indicated by our interviewees (and supported by **F6:** Lack of common time), things such as time-zones, working hours, weekends and holidays would most likely need to be taken into consideration when calculating the statistics mentioned for **FT19, FT20** and "reply speed to node category" part of **FT25**, in order to improve the accuracy of this information. Allowing users to connect nodes (based on email addresses) together using **FT8**: **Combining nodes by email address** should also improve accuracy, not just for that user's statistics but also for the statistics of any other users previously connected to the nodes that were combined. Finally, the time span filter mentioned in **FT34: Filtering visualization** can be applied in order to improve accuracy (by making sure the currently active timespan is relevant).

F9: Low Prioritized interaction and **R7:** Creating awareness indicates that by creating awareness of the status of people's recent email conversations, you can help people understand the priorities and status of emails so that they are aware of its importance to the people they interact with, which is why we designed **FT22: Recent Conversation Priority** and **FT24: Recent Conversation Status Icons** in addition to **FT20**. However, our suggested features related to recent conversations such as **FT20, FT22,** and **FT24** are quite simplistic in order to avoid privacy issues and display of "unsuitable" email topics, as well as to reduce the amount of user input needed for the features to work (which is why status and priority are to be automatically generated).

5.2. Validity
The aspects of validity we discuss are based on the four types of validity described by Runeson & Höst [13], and the dangers Hevner et al. [2] warn of in regards to Design Science Research projects. Specifically we have used that information in our effort to determine potential threats to our study's validity, as well as assess if we have avoided them.

Possible threat: Interviewees not understanding what we show/ask
In order to avoid this threat, we used short presentations of our topic and short descriptions of the actual mockups before we actually started the interviews, ensuring that the interviewee gained some basic understanding about what we were doing and what we were showing them.

Possible threat: Mockups not accurate representation of future prototype
The mockups we have produced as part our contributions do not represent the "final" look on the system (based on what we know we can do with the tools we plan to use for prototyping). This is due to how the various layout algorithms encountered through our literature review and the technology survey are not part of these static mockups, since there was no efficient way of doing this.

Possible threat: Too much focus on technology
In [2] there are descriptions of several common threats to validity in design science research papers. One such danger is how researchers can overemphasize on the technological artifacts

over all other aspects of the research, potentially resulting in well-designed artifacts that are useless in real organizational settings. We avoided this danger by leaving the implementation of our prototype outside the scope of this research paper, using guidelines, mockups and scenarios to represent our design instead.

Possible threat: Lack of similar tools
Another danger Hevner et al. [2] warns about is how when the existing knowledge base is lacking, designers may have to rely solely on intuition, experience, and trial-and-error methods. By using mockups rather than an actual prototype for showing our designs however we somewhat circumvented this, but this danger is still a large risk for our future work when we start implementing the prototype (since we do not have any examples of other email network visualization tools used for interaction speed improvement).

Possible threat: Artifacts tailored to specific research setting
Yet another danger mentioned in [2] is how a design artifact on a single project may not generalize to different environments. This was also a motivation for us to focus on general features and design guidelines based around identified issues, and mockups giving examples of how the visualization could look like, rather than focusing on a specific implementation.

Possible threat: Out-of-date results
Through our technology survey, we have managed to establish a general design/architecture for our future prototype, including decisions on the languages, frameworks and tools that we plan to use. However, as stated by Hevner et al. [2], design-science research is perishable, due to the rapid advancement rate of technology. This was yet another reason for us to focus on mockups and guidelines rather than on an actual implementation.

Possible threat: Too theoretical
As we have mentioned in regards to other potential threats, this study is very much focused on creating a theoretical design rather than making a prototype, meaning that we have no physical "proof" of our concepts. However, the fact that our mockups have led to us getting both feedback that led to improvements and acceptance on parts of our designs indicate that we are on the right track. The fact that we have identified tools/API:s that we can use to implement our features also adds "realism" to our designs.

Possible threat: Unclear research process
One common threat to validity is if a study is replicable or not. However, since we have provided detailed descriptions of our methods and of the structure of this project (in compliance with the contributions that Hevner et al. [2] state that a design science research paper should provide), we have eliminated this threat.

Possible threat: Researcher bias
In regards to our research setting, our familiarity with it could be cause for concerns on researcher bias, especially in relation to the interviews with CSE employees and students. We feel however that this has not been an issue, since the fact that we recorded the interviews has allowed us to reevaluate the responses several times. For iteration 2 interviews, we also made sure to actively ask about what aspects of our ideas/mockups the interviewee's didn't like, in order to identify areas where we could make improvements. More general bias issues have been alleviated somewhat due to the fact that both researchers participated in all parts of this study's process (Runeson & Höst [13] claim that such investigator triangulation helps to improve validity).

5.3. Future work

The main future use of our study that we are considering is to follow through with our preparations for iteration 3, and start developing an initial prototype based on our findings and contributions. Another option would be to extend the research to other communication networks besides email, or to replicate the study at another organization in order to further validate the generality of our contributions. During this study we also discovered another research angle that we are considering for generating updates to our design. Specifically, we have concerns on how to ensure that the visualization tool we develop gets a large enough userbase to be relevant (this is also important for the features that rely somewhat on user manipulation). We believe that we could generate such features focused on attracting more users by looking into the research area of Gamification. Another interesting suggestion for future work would be to look into the other aspect of SNA (datamining) in relation to this study. Specifically, it would be useful to design a datamining system for real time email reading in order to feed fresh data to the visualization.

For anyone that wishes to replicate our study or conduct a similar one, we would recommend that they take the legal and ethical aspects of monitoring communication and displaying personal information into account from the very beginning of their design (since this is something we failed to do initially, which later on led us to revising and simplifying certain features).
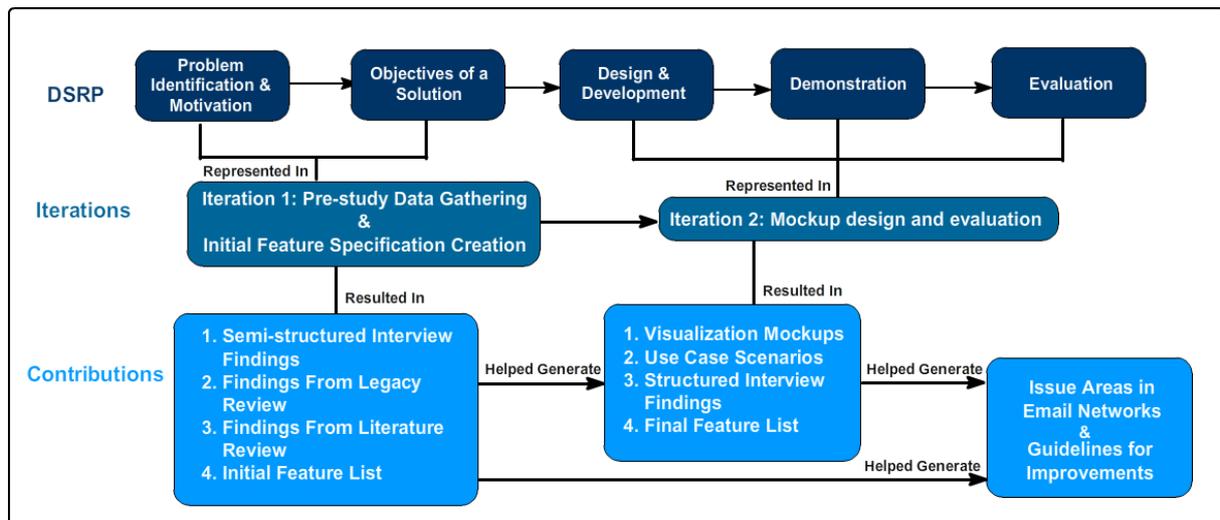
6. CONCLUSION

Figure 6a. *Summary of our study*

In our research to find ways to improve interaction speed in organizational email networks through designing a visualization application, we have conducted a study consisting of two iterations that were planned using **DSRP** [6]. We have gathered data through the use of literature review, technology survey, and interviews, and analyzed it with the purpose of extracting feature ideas for our visualization design. This resulted in us producing a feature list and a basic architecture for the visualization (along with suggestions for what components to use as part of this architecture). We also created mockups based on those artifacts, together with scenarios that clarify what these mockups represent. We have also provided visualization guidelines for handling two major identified issue areas regarding interaction speed within organizational email networks (**Determining who to contact** and **Detecting presence of**

interaction speed slowdown factors and effects). Through these contributions we have answered our research questions of:

- ***RQ1:*** *What issues cause interaction speed slowdown in email networks?* (through the identified issue areas of "Determining who to contact" and "Detecting presence of interaction speed slowdown factors and effects")
- ***RQ2:*** *How can we design a visual solution for monitoring and improving interaction speed within email networks?* (through the feature list, architecture, scenarios, mockups and guidelines)

By answering our research questions we have also contributed to filling the knowledge gap of interaction speed improvement through visualization of email networks.

Based on the issue areas we identified, we believe that they can be applied to visualizations of other types of informal communication networks (rather than email only). An interesting future study would then be to test this hypothesis. It would also be useful to replicate or to conduct a similar study of the email network at another research setting, in order to further validate the usefulness of our designs and guidelines.

7. REFERENCES

[1]. A. Martini, L. Pareto and J. Bosch. "Improving Businesses Success by Managing Interactions among Agile Teams in Large Organizations," presented at the 4th Int. Conf. Software Business, Potsdam, Germany, 2013.

[2]. A. Hevner, S. March, J. Park and S. Ram. (2004, March). "Design Science in Information Systems Research." *MIS Quarterly*, [On-line]. 28(1), pp. 75-105.
Available: http://dl.acm.org/citation.cfm?id=2017217 [May 23, 2013].

[3]. B. Kerr. "Thread Arcs: an email thread visualization," in *Proc. INFOVIS,* 2003, pp. 211-218.

[4]. B. Christian, G. Alex, D. Prem, G. Michael, and S. Anand. "Mining email social networks," in *Proc. MSR,* 2006, pp. 137-143.

[5]. K.W. Brodlie, D.A. Duce, J.R. Gallop, J.P.R.B. Walton and J.D. Wood. (2004, July). "Distributed and Collaborative Visualization." *Computer Graphics Forum*. [On-line]. 23(2), pp. 223-251.
Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2004.00754.x/abstract
[May 23, 2013].

[6]. K. Peffers, T. Tuunanen, C. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge. "The Design Science Research Process: A Model for Producing and Presenting Information Systems Research." in *Proc. DESRIST*, 2006, pp. 83-106.

[7]. M.F. Schwartz and D.C.M. Wood. (1993, Aug.). "Discovering shared interests using graph analysis." *Communications of the ACM*. [On-line]. 36(8), pp. 78 – 89.
Available: http://dl.acm.org/citation.cfm?id=163402 [May 23, 2013].

[8]. M. Jamali and H. Abolhassani. "Different Aspects of Social Network Analysis," in *Proc. WI,* 2006, pp. 66-72.

[9]. M. Myers and M. Newman. (2007, Jan.). "The qualitative interview in IS research: Examining the craft." *Information and Organization*. [On-line]. 17(1), pp. 2-26.
Available: http://www.sciencedirect.com/science/article/pii/S1471772706000352 [May 23, 2013].

[10]. M. Fris, M. Nilsson and V. Sollerhed. "Real-Time Social Network Visualisation." B.I thesis, University of Gothenburg, Sweden, 2011.

[11]. M.K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. (2011, March). "Action design research." *MIS Quarterly*. 35(1), pp. 37-56. Available: http://bada.hb.se/handle/2320/9888 [May 23, 2013].

[12]. M.E.J. Newman, S. Forrest and J. Balthrop. (2002,). "Email networks and the spread of computer viruses." Physical Review E. [On-line]. DOI: 10.1103/PhysRevE.66.035101.
  Available: http://link.aps.org/doi/10.1103/PhysRevE.66.035101 [May 23, 2013].

[13]. P. Runeson and M. Höst. (2009, April). "Guidelines for conducting and reporting case study research in software engineering." *Empirical Software Engineering*. [On-line]. 14(2), pp. 131-164. Available: http://link.springer.com/article/10.1007/s10664-008-9102-8 [May 23, 2013].

[14]. R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas .(2006, Dec.). "The real communication network behind the formal chart: Community structure in organizations." *Journal of Economic Behavior & Organization*. [On-line]. 61(4), pp. 653-667.
  Available: http://ideas.repec.org/a/eee/jeborg/v61y2006i4p653-667.html [May 23, 2013].

[15]. S. Hove and B. Anda. "Experiences from Conducting Semi-Structured Interviews in Empirical Software Engineering Research." in *Proc. METRICS,* 2005, pp. 10-23

[16]. S. Huang and M. Anton. "Conference Networking Recommendations based on Past Online Activity." B.S thesis, University of Gothenburg/Chalmers University of Technology, Sweden, 2012

[17]. X. Fu, S.H. Hong, N.S. Nikolov, Shen X., Y. Wu, and K. Xu. "Visualization and analysis of email networks," in *Proc. APVIS,* 2007, pp. 1-8

APPENDIX A: DEFINITIONS

Here we present the definitions of each factor, effect and recommendation, as found in [1].

Factors & Recommendations
F1. Knowledge unavailability
If a team doesn't have all the knowledge to develop a feature independently, they will try to interact with an expert outside the team, creating interactions. The team may alternatively decide to make assumptions on the answers that lead to redo most of the work.
Recommendation R1: experts made available for the teams covering critical knowledge (the most requested one) would avoid unnecessary interactions.

F2. Expert's reputation
If an employee has a high reputation of having a specific knowledge, the person will be contacted often. Reputation is not only based on the real knowledge of an employee, but rather on his or her social reputation.

F3. Unclear requirements
This relates to two interaction problems: the long waiting time before the team is able to receive the specification, or the continuous interaction for clarification of the requirements afterwards.
Recommendation R2: ASD relies on more volatile requirements and their continuous adjustment: this has to be supported by the (part-time) presence of an architect ("architecture customer") and a product manager ("product owner") on site; otherwise the inter-team interactions would be substantial and not effective because external with respect to the team.

F4. Unexpected Feature Dependencies
Two features may be designed to interact with each other through APIs or through a component. In some cases, dependencies pop up unexpected, e.g. due to indirect (software) interactions or because of socio-technical reasons.
Recommendation R3: In this case, as in other kinds of team, a bridgehead between the two teams would help such coordination.

F5. No co-location
Large organizations are forced to spread teams in space. According to our interviews, even the distance of one floor makes them distributed, with consequent delays and lack communication and commitment.

Recommendation R4: The interviews suggest that the teams that have to interact more intensely should be located closer. It can be considered as another level of co-location with respect to intra-team co-location.

F6. Lack of common time
Teams may need to synchronize in meetings, which require common available time. If a team decides to not allocate time for interaction or the allocated time-slots don't match, there is a lack of communication or long waiting times. Causes may be the different locations, different time zones (or with different slots of working hours), calendar interferences or low prioritized interaction.

Recommendation R5: some agile practices, such as SCRUM, include support for meetings between SCRUM masters. However, other kinds of programmed available time could be considered. **R6:** Also shared calendars would help finding better alignment.

F7. Mismatch of team's styles of communication
Different teams may have different "styles" of communication, which may cause delays: e.g. one team mainly uses e-mails and doesn't want to meet in person. The Agile culture of letting teams having their customized processes somehow encourages this mismatch. **Recommendation**: Good practices would be of help on how to maintain inter-team interfaces, for example with bridgeheads, employees that have a strong influence in more than one team (**R3**). Also **R2** may be applied, if architecture or management teams are involved.

F8. Slow resource indexing
When a member of a team needs to interact, he or she needs to find the correct person or team to interact with. The time spent on such activity may be long and therefore delaying.

F9. Low prioritized interaction
Once an interaction is needed, the involved parts (single employees or whole teams) have to prioritize the interaction as an on-going task. If the interaction is considered as "low priority", the team will delay tasks and communication, hindering the other team(s) involved. **Recommendation R7**: Tools for creating awareness would help understanding the overall situation of the involved teams. Again, the presence of people also connected to other teams would enhance commitment (**R3**, **R2**).

F10. Inter-personal conflicts
Two employees in different teams (or even the whole teams) may consider each other "enemies" (for personal or political reasons). Interactions between these employees may be strongly hindered by delays and corrupted information. Again, a work environment strongly built on social interactions may amplify this factor. Recommendations for this factor have to be rooted in psychology and theory of organization.

Effects
E1. Long waiting time to communicate
A team has to wait before communicating with other ones.

E2. Long waiting time for value
As the previous one, but the team is waiting for the realization time needed from the other team to deliver the value.

E3. Intense communication
Each instance of inter-team communications requires long time.

E4. Corrupted communication
The information received by the team is insufficient to deliver the requested value. This, in turn, may cause intense communication or high interaction frequency.

E5. High interaction frequency
The number of interactions between two teams is too high. A special instance of this effect occurs when a member in the team is continuously consulted for his or her knowledge by many other teams.

E6. High task frequency
A single interaction may require many tasks to be carried out in order to deliver the value.

E7. Heavy interaction tasks
The time is long because of the large amount of time required for carrying out the task to deliver the value.

E8. Corrupted value
A (sub-) value has to be delivered to complete the interaction. However, the received value doesn't satisfy the need that started the interaction.

APPENDIX B: SCENARIOS & MOCKUPS

During iteration 2, we conducted a short legacy review in order to establish what tool to use when producing our mockups. Initial searching led us to try out different wireframe mockup tools, due to their ease of use. A wireframe mockup tool lets you simply drag and drop the elements you want to use in your mockup, to the position that you want to have them in. The main drawback however is that it is hard to emulate the algorithm based layouts available via the visualization tools, such as force-directed and hierarchical clustering. It is also hard to emulate some of the available interactivity/animation features from those tools. While there are many wireframe mockup tools available, we focused on finding one that would easily let us modify size, shape, colour and text for the elements we use. The one we ended up using (Mockflow) provides these features, along with allowing us to collaborate online within the same mockup project. The web based nature of the tool also let us access and modify our mockups from any computer, which was helpful due to how we did not always work on the project from the same location.

Below, we show the latest versions of these mockups, in order to give a visual representation of our design choices. We have also included use case scenarios that explain what is happening in these images. In these scenarios, we also point out the related features.

Scenario 1: Combining Email Addresses
Description
This scenario shows the process of a user who logs in, goes to their user page and uses account management features to link a node based on another email address to the account they are logged in on. The purpose of this is to get more accurate interaction statistics for that person that for some reason use more than one email address. In the mockups accompanying this scenario, these features are represented visually:

- **FT3: Use of Existing Authentication Systems**
- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT13: Information Panel**
- **FT14: Selection Emphasis**
- **FT15: Quick-select Self**
- **FT16: People Panel**
- **FT17: Hideable Components**
- **FT25: Additional Information on Node Selection**
- **FT33: User Page Management**
- **FT8: Combining nodes by email address**

Priority
Medium

Preconditions
User must have two email addresses that are both represented in the visualization.

Postconditions
The two nodes representing these email addresses have been combined, and both emails can now be used to log in on the same account.

Actors
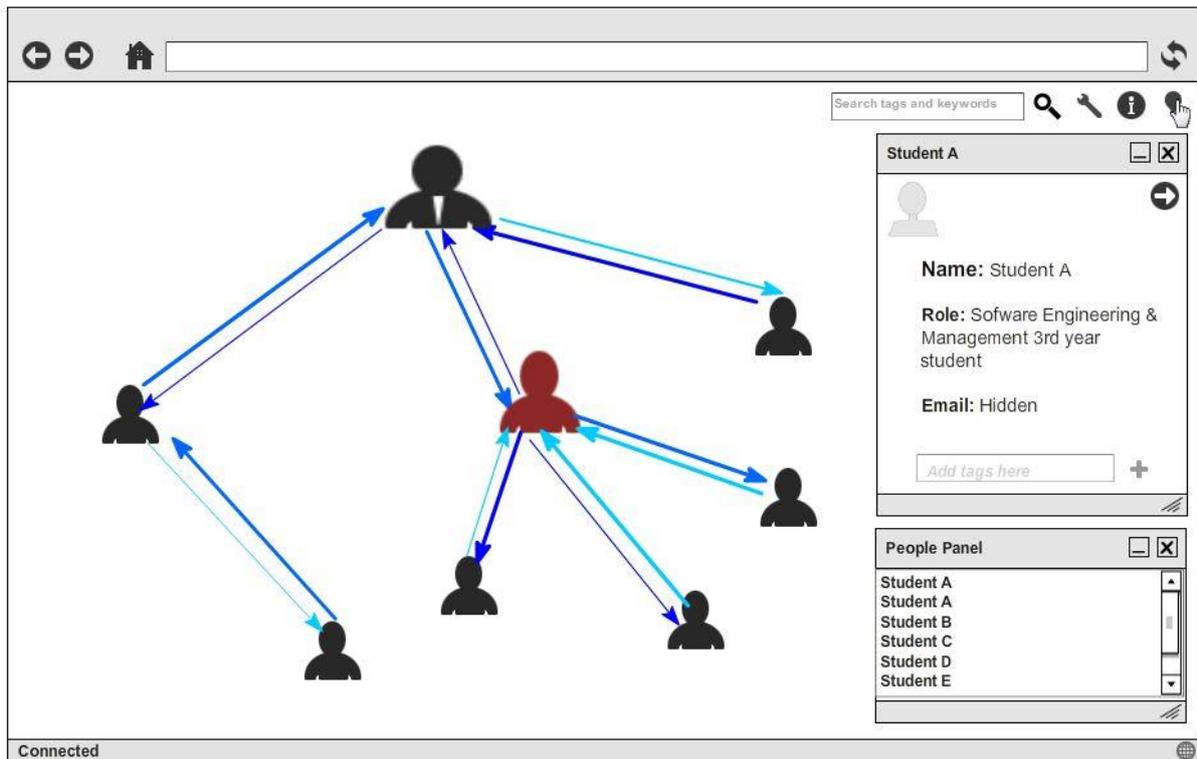The user that owns of both these nodes.

Basic Flow
1. User logs in via GU authentication system.

B1. Authentication system example

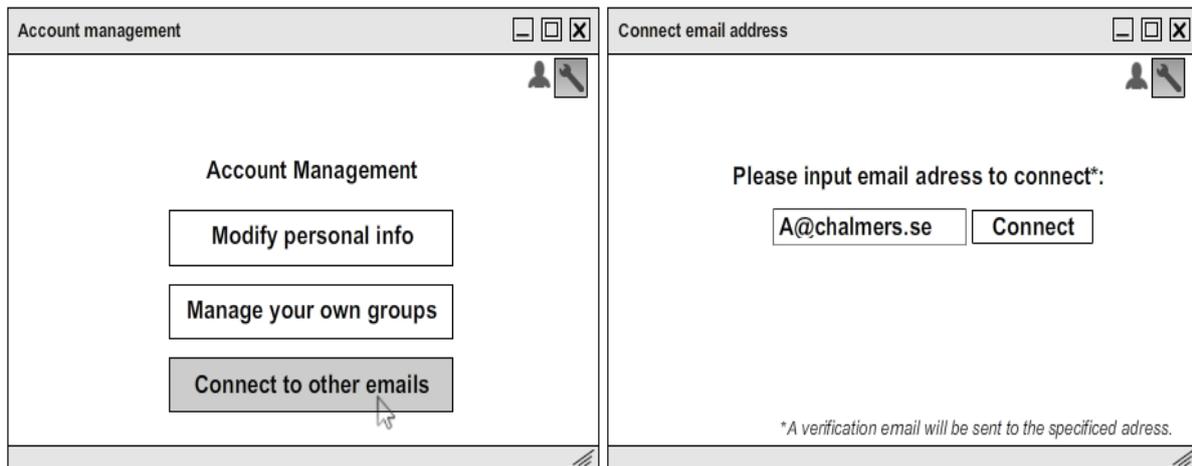2. User uses shortcut feature to find and view their own node.

B2: Main view mockup

3. User goes to own page.
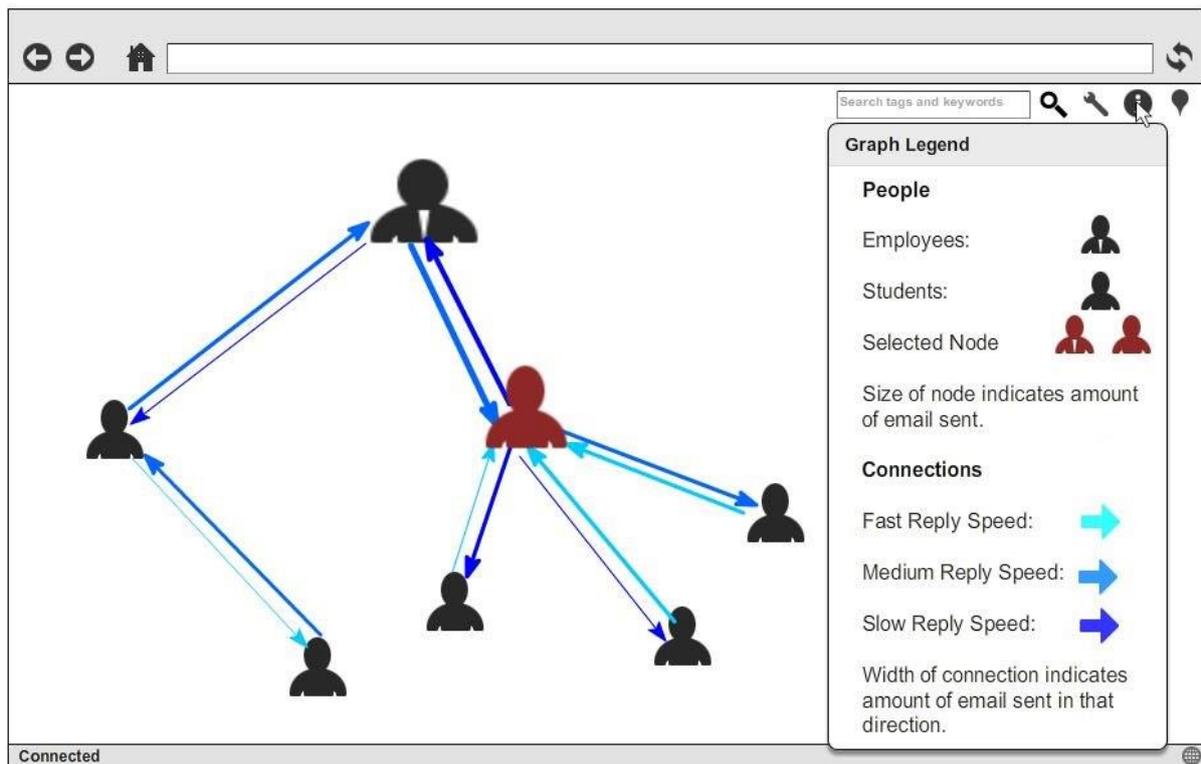
4. User goes to account management page.

5. User attempts to connect the node representing their Chalmers email to the account they are currently logged in on.

B3. Email connection mockup

6. Confirmation mail sent to Chalmers email is activated by user.

7. User goes back to graph and sees the changes based on two nodes combining.

B4. Merge result mockup

8. User can now use Chalmers and GU authentication systems to access the same node.

Alternative flow
6a. Confirmation mail sent to Chalmers email is not activated by user.
- 6a1. The nodes remain separate.
- 6a2. After one week, the confirmation mail is expired.

Scenario 2: Finding Person from Recent Conversation
Description
This scenario shows the process of how a user can attempt to find another user that they were recently emailed by, in order to see more information on this person. The purpose of this can be linked to gaining knowledge on that other person in general (such as what their role is), and how they can be found if other forms of communication besides email is required. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT12: "Hidden" Detailed Information**
- **FT13: Information Panel**
- **FT14: Selection Emphasis**
- **FT15: Quick-select Self**
- **FT16: People Panel**
- **FT17: Hideable Components**

- **T19: Overtime Interaction Statistics on Connection**
- **FT21: Tags for Recent Conversations**
- **FT22: Recent Conversation Priority**
- **FT24: Recent Conversation Status Icons**
- **FT29: User Page**
- **FT30: Free-text for adding information**
- **FT31: Location Information**
- **FT32: Shared Calendars**
- **FT35: Highlighting Recent Conversation**
- **FT36: Partial Filter/Highlight**

Priority
High

Preconditions
Must be logged in.
Must have been recently contacted by the person you are looking for.
The person you are looking for must be part of the visualization.
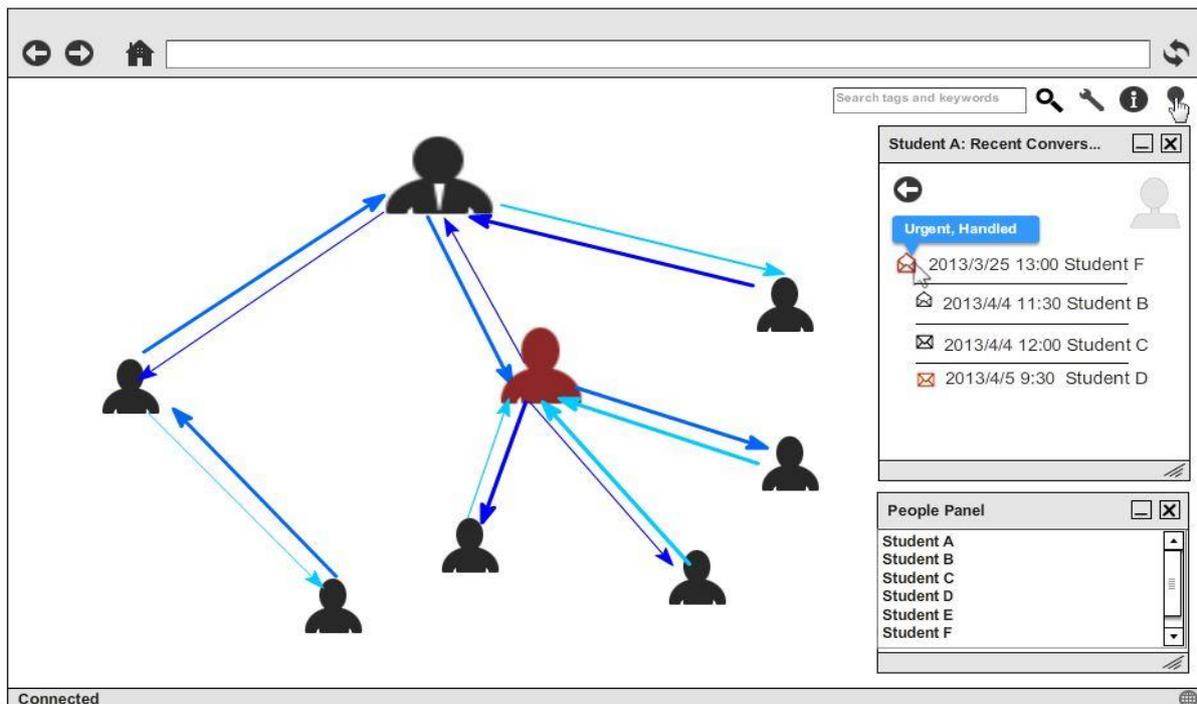
Postconditions
Additional information on the person has been found, and the user now has the option to contact him/her through other means than email (if required).

Actors
The user who was contacted.

Basic Flow
1. User selects own node and switches to recent conversation view in information panel.
2. User hovers on the icon for the first conversation.

B5. Recent conversations mockup

3. User selects the first conversation due to its tag.

4. User selects a person who was in this conversation via the people panel.
5. User hovers over the (highlighted) edge connecting this person to him/her.
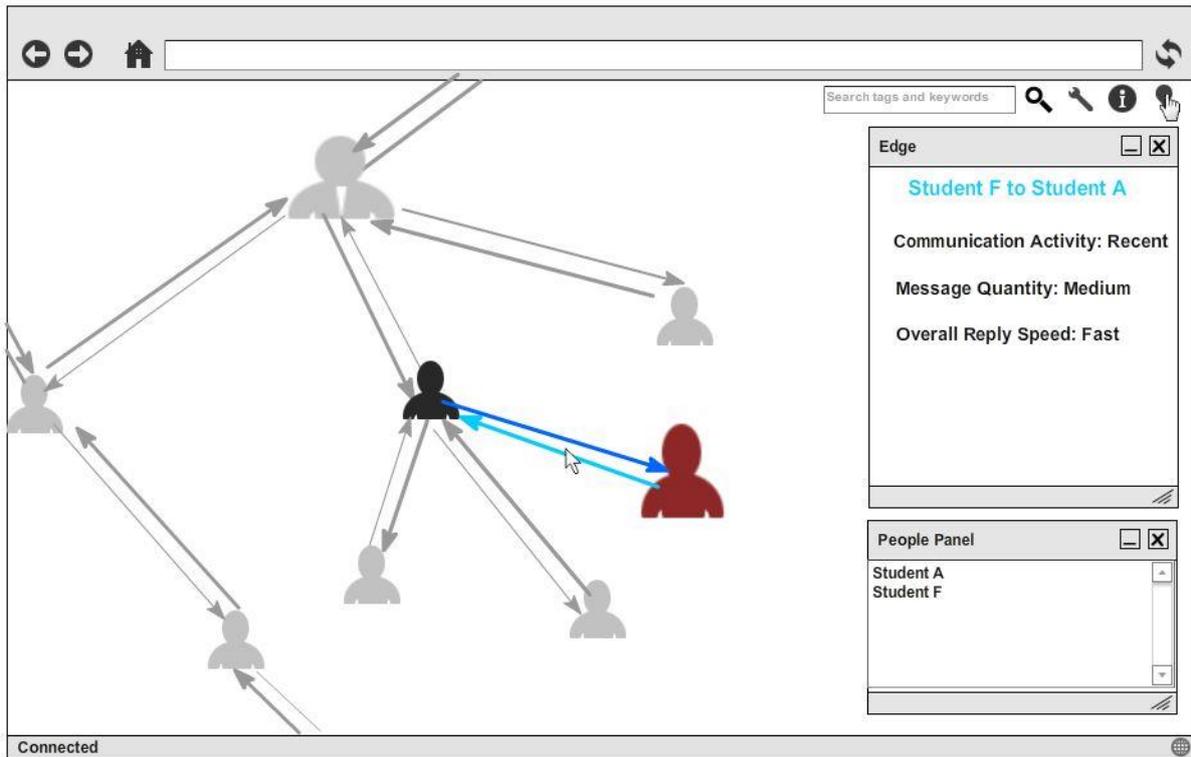
B6.Overtime interaction statistics mockup

6. User selects that person's node, and sees based on picture that it is the right person.
7. User gets the necessary information from looking at the information panel while having this node selected.

Alternative Flow
6a. User does not know what the person looks like/is not sure if this is the right person.
- 6a1. User has to go to the person's user page for more information.
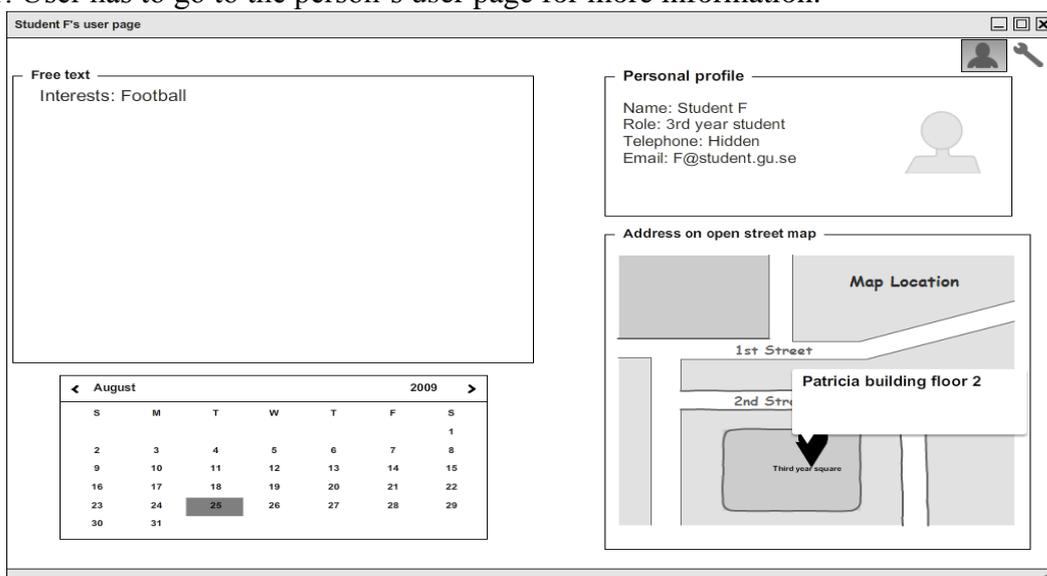
B7. Userpage mockup

Scenario 3: Adding tag to user's own node

Description:
This scenario shows how a user can try to add tags to his/her own node in the visualization. The purpose here is again related to things that make finding the right person easier. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT25: Additional Information on Node Selection**
- **FT26: Node Tag System**

Priority:
High

Preconditions
The user must already be logged in.
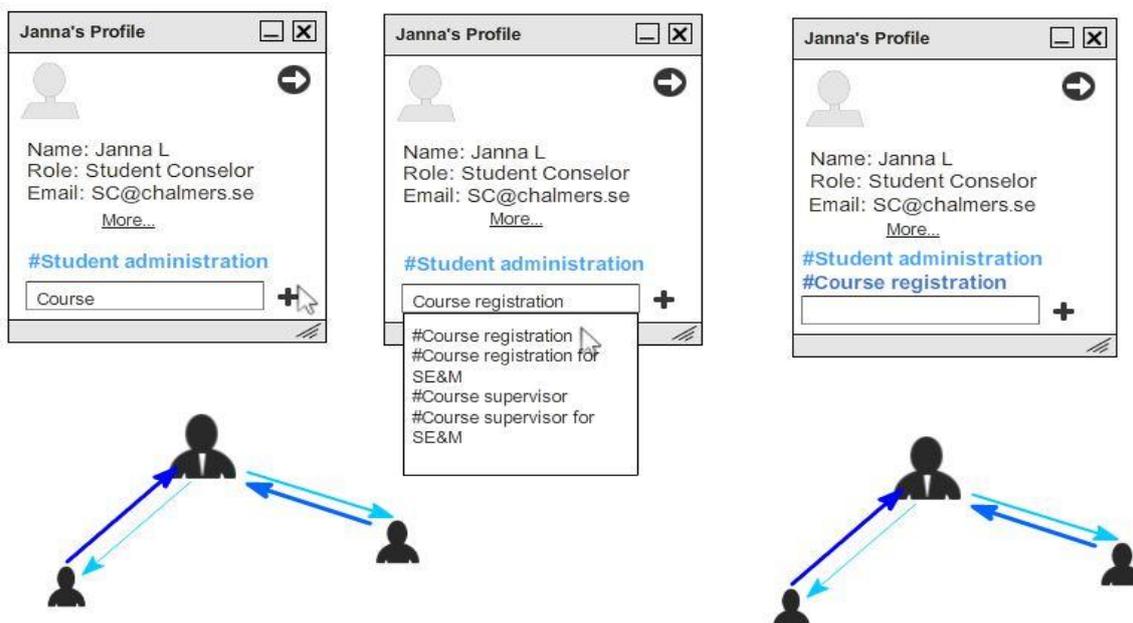
Postconditions
The user has gained tag(s) on their own node.

Actors: User, Administrator

Basic flow:
1. User selects his/her own node.
2. System represents the node information.
3. User writes down keywords.
4. System generates predetermined tags.
5. User chooses one of the predetermined tags.
6. User adds tag to his/her own node.

B8. Node tag system mockup

Alternative flow 1:
5a. User creates a new tag other than the predetermined tags.
- 5a1. Administrator accepts the new tag.
- 5a2. User adds tag to his/her own node.

Alternative flow 2:
5b. User creates a new tag other than the predetermined tags.
- 5b1. Administrator rejects the new tag.
- 5b2. User fails to add tag.

Scenario 4: Adding tag to others' node

Description:
This scenario shows how a user can try and add tags to other users' nodes in in the visualization. The purpose of this is to allow users to tag related people (such as coworkers) who are not frequent users of the visualization themselves (of course with restrictions, as mentioned in **FT27: Node Tagging Restrictions**). This scenario is visually based around the mockup shown in the previous scenario (*B8*).

Priority:
High

Preconditions
User is already logged in.

Postconditions
The node targeted by the user has gained tag(s).

Actors: User, The node owner, Administrator

Basic flow:
1. User selects one node other than his/her own.
2. System presents the node information.
3. User writes down keywords.
4. System generates predetermined tags.
5. User chooses one of the predetermined tags.
6. The node owner accepts the tag from user.
7. User adds tag to that node.

Alternative flow 1:
5a. User creates a new tag other than the predetermined tags.
- 5a1. Administrator accepts the new tag.
- 5a2. The node owner accepts the new tag from user.
- 5a3. User adds tag to that node.

Alternative flow 2:
5b. User creates a new tag other than the predetermined tags.
-5b1. Administrator rejects the new tag.
-5b2. User fails to add tag.

Alternative flow 3:
6a. The node owner rejects the tag from user.
- 6a1. User fails to add tag.

Scenario 5: Searching with keywords

Description:
This scenario shows how a user can attempt to find relevant nodes through filtering the visualization by searching for certain keywords. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT17: Hideable Components**
- **FT34: Filtering Visualization**
- **FT36: Partial Filter/Highlight**

Priority:
High

Preconditions
User is already logged in.

Postconditions
The user's current view has been altered in order to highlight elements relevant to the tag(s) he/she has searched for.
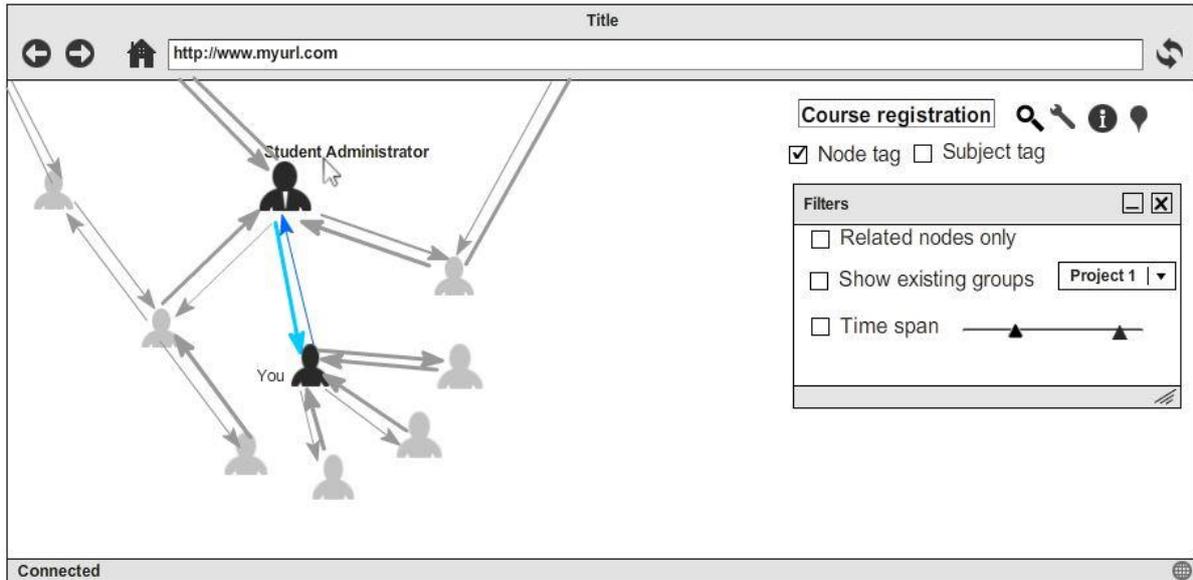
Actors:
User

Basic flow:
1. User inputs keywords in search bar.
2. User selects node tag filter.
3. User starts searching.

4. System highlights the search result in the visualization.

B9. Node tag search mockup

Alternative flow 1:
2a. User selects subject tag filter.
- 2a1. User starts searching.
- 2a2. System highlights the search result in the visualization. (Alternative flow 2)
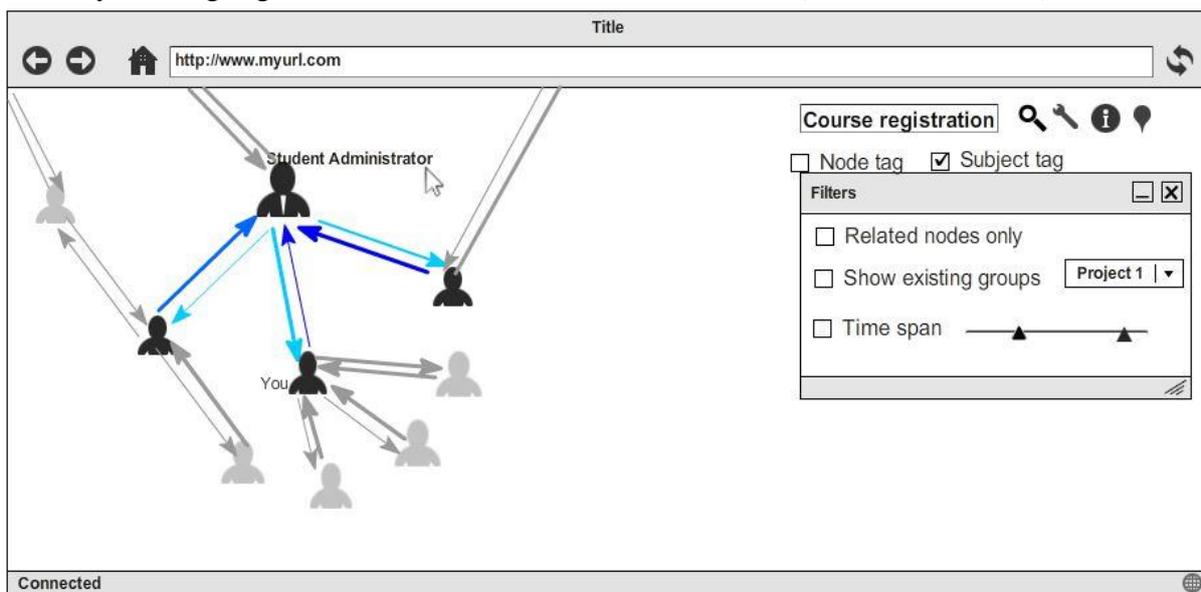
B10. Subject tag search mockup

Alternative flow 2:
4a. System returns "no result found."

Scenario 6: Creating custom group
Description:
This scenario shows how a user can attempt to create a custom group (for grouping certain nodes together). In the mockups accompanying this scenario, these features are represented visually:

- **FT11: Icon Interface**
- **FT28: Custom Groups**

Priority:
Low
Preconditions
User must be logged in

Postconditions
The group creator and invited users are now part of the same custom group (which means that they will be highlighted when one filters the visualization based on this group).

Actors:
User creating the group, admin checking the group name and any other users that are invited.

Basic flow:
1. User selects settings icon (cog).
2. User types in potential group name that doesn't exist.
3. Admin accepts the group name.
4. User invites other users to the group.
5. The users accept.

B11. Custom group creation mockup

Alternative flow 1:
2a. If the name already exists, return to 2.

Alternative flow 2:
4a. Invitation declined.

Scenario 7: Using custom group filter
Description:
This scenario shows how the visualization can be filtered based on custom groups. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**

- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT17: Hideable Components**
- **FT34: Filtering Visualization**
- **FT36: Partial Filter/Highlight**

Priority:
High

Preconditions
User is already logged in.

Postconditions
The user's current view has been altered in order to highlight the members of the custom group chosen in the filter settings.

Actors:
User

Basic flow:
1. User selects filter "Show existing groups".
2. User selects one group from the scroll down list.
3. System represents the customized visualization.
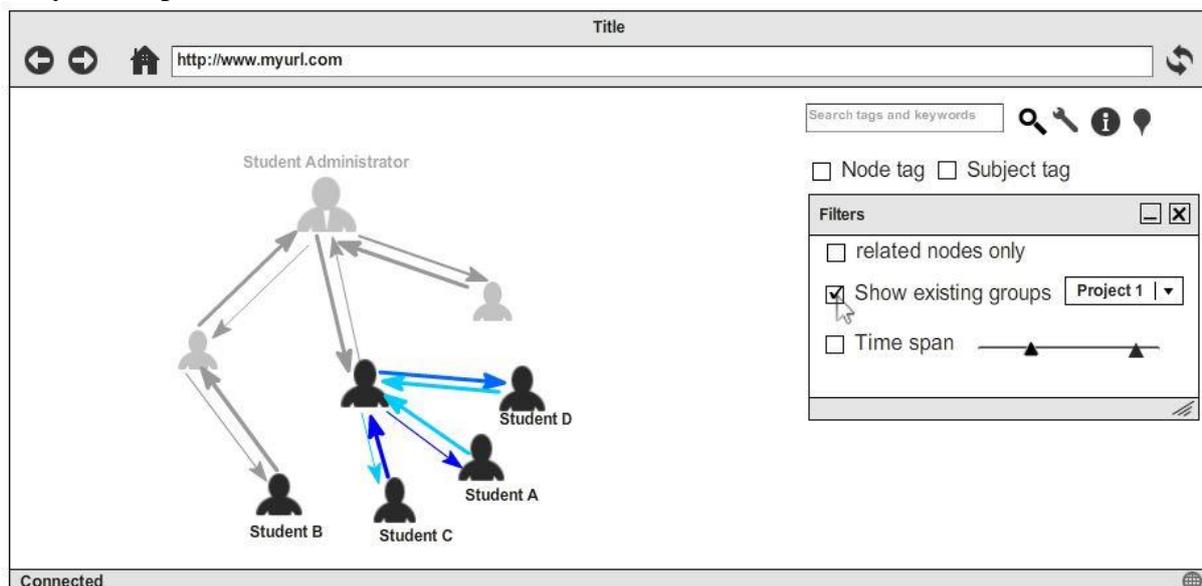
B12. Custom group filter mockup

Alternative flow 1:
2a. User selects one node in the visualization.
- 2a1. User selects one group from the scroll down list. (has Alternative flow 3)
- 2a2. System represents the customized visualization.

Alternative flow 2:
2b. No existing groups in the scroll down list.
- 2b1. User fails to use this filter.

Scenario 8: Using node relation filter

Description:

This scenario shows how the visualization can be filtered based on nodes related to the currently selected node. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT17: Hideable Components**
- **FT34: Filtering Visualization**
- **FT36: Partial Filter/Highlight**

Priority:

High

Preconditions

User is already logged in.

Postconditions

The user's current view has been altered in order to highlight the nodes that have communicated with his/her currently selected node.

Actors:

User

Basic flow

1. User selects own node.
2. User selects filter "show related nodes only".
3. System represents the customized visualization.
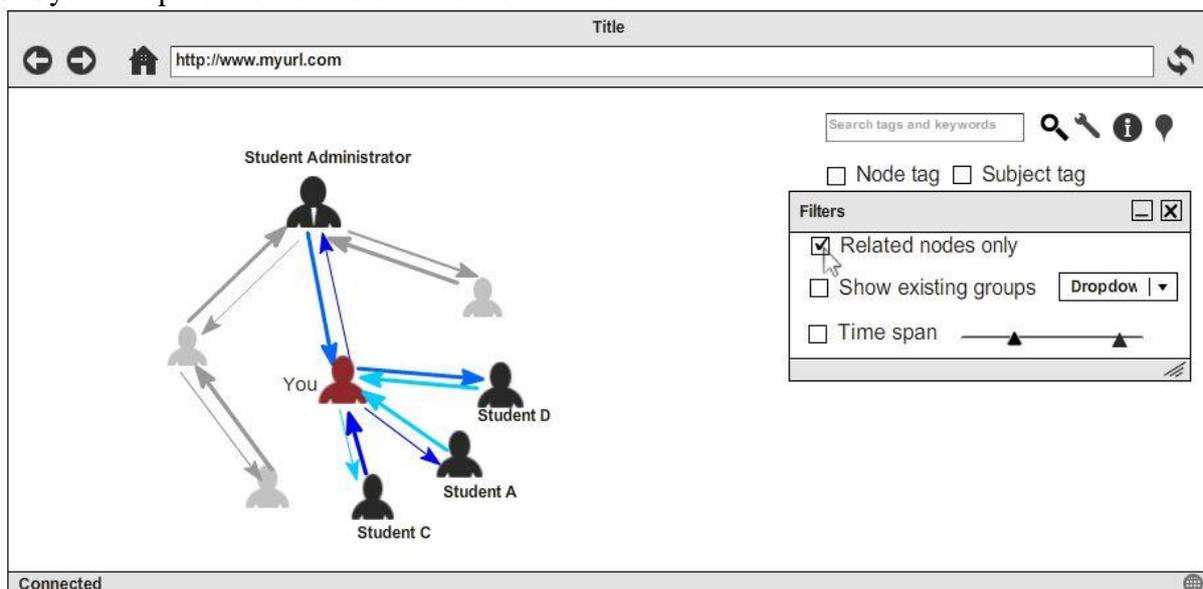
B13. Node relations filter mockup

Scenario 9: Using timespan filter

Description:

This scenario shows how the visualization is affected when the timespan filter is adjusted. In the mockups accompanying this scenario, these features are represented visually:

- **FT4: Show Communication in Both Directions**
- **FT5: Using Colours & Shapes to represent information**
- **FT6: Coulourblind-adjusted Colour Scheme**
- **FT7: Icon variation for major roles**
- **FT11: Icon Interface**
- **FT17: Hideable Components**
- **FT34: Filtering Visualization**
- **FT36: Partial Filter/Highlight**

Priority:

High

Preconditions

User is already logged in.

Postconditions

The change to the current timespan has caused relations and interaction speed statistics to be recalculated, which in turn has altered the user's view.

Actors:

User

Basic flow:

1. User selects filter "Time span".
2. User chooses one date as start point within the time span.
3. User chooses another date as end point within the time span.
4. System represents the customized visualization.

B14 Timespan filter mockup

APPENDIX C: DSRP

Peffers et al:s DSRP [6] divides research projects into six steps. Below these steps are explained and linked to this project.

1. **Problem identification and motivation**
   This activity involves defining the specific research problem and why it needs to be solved, using knowledge of the state of the problem and the importance of its solution. This was achieved through discussions with our main stakeholder at the start this project, as well as a review of the current state of related literature (represented in the two study areas mentioned in our theoretical framework).

2. **Objectives of a solution**
   This activity involves extracting the objectives of a solution from the problem definition. This was achieved by analyzing the literature found through our review, as well as the responses of interviews with several CSE employees.

3. **Design and development**
   This activity involves determining the artifacts' desired functionality and its architecture, and then creating the actual artifact(s). Design Science Research artifacts include constructs, models, methods, or instantiations [2]. This was achieved through writing a feature list and defining a basic architecture based on the collected data both from the mentioned interviews and literature analysis, as well as looking at several visualization tools. We then created mockups based on this feature list and architecture. In the future, this activity may be touched upon again, with the purpose of designing and developing a functional prototype based on mockup feedback.

4. **Demonstration**
   This activity involves demonstrating the potential of the artifact to solve the problem. This was achieved via mockup demonstrations to people within the research setting (which also included additional interviews based around these mockups).

5. **Evaluation**
   This activity involves comparing the objectives of a solution to actual observed results from use of the artifact in the demonstration. Based on the five design evaluation techniques available on page 12 in [2], the technique we found most suitable evaluation for this study was the descriptive evaluation, due to its qualitative nature. Descriptive evaluation is achieved by using information from relevant research to build argumentation for supporting the artifact's utility [2] (which we do through our guidelines in our Discussion). Descriptive evaluation also includes constructing scenarios that highlight the functionality of the artifact(s). We provide such scenarios in Appendix B: Scenarios & Mockups, where they are used to explain what can be seen in our mockups.

6. **Communication**
   This activity involves communicating the problem and its importance, as well as the artifacts' utility and effectiveness, to relevant audiences. This activity has been handled through the writing this paper as well as project presentations on this topic.

APPENDIX D: DETAILED FEATURE EXTRACTION FROM FACTORS, EFFECTS & RECOMMENDATIONS

Parts of our interview results such as responses like "Lacking knowledge of who to contact" and "accurate information of how to handle the request" can be linked to **F1 Knowledge unavailability**. One example of this is how if an email receiver does not have all the knowledge to resolve or answer a request; they tend to have more interactions with the people who sent the request in the first place. In other cases, if an email sender does not provide enough information about himself/herself, or if the definition of the request is lacking or information needed is missing, it will also result in additional interactions. By visualizing the roles of individuals in email networks, it may be possible to decrease the risk of this knowledge unavailability (since such information should help indicate who has what knowledge). This could possibly also be handled in part by having users add more information about themselves manually. Additionally, allowing the user to filter the visualization based on certain keywords and categories should help with finding people with certain knowledge. These measures would allow people to more easily identify multiple people with certain roles, so that they can spread out some of their requests in order to reduce **E6 High task frequency** and **E7 Heavy interaction tasks**. If users are able to manually group their nodes in the visualization with others' nodes, their group memberships could also help indicate who is working within the same area. The recipient of requests via email could of course also use the visualization to personally identify others that could help them with the request.

By editing customized information such as role, available time etc. for each node, and by having that information available for all users, (part of) the missing information that causes **F3 Unclear requirements** should also be findable via the visualization. This information should also help reduce the effects of **F4 Unexpected Feature Dependencies** and **F8 Slow resource indexing**, since it should increase peoples' awareness of each other's responsibilities within the organization, making it easier for people to find who to cooperate with. Additionally, it can also help reduce the risk of **E4 Corrupted communication** and **E8 Corrupted value**. These previously mentioned methods for using the visualization to create awareness of each person's area of expertise and/or interest can also be linked to **R1 Make experts available**.

Another link we have discovered is the possibility of viewing the effects of **F2 Expert's reputation** in our visualization. Specifically, parallels can be drawn between employees' positions and how much interaction they receive. For example, as part of our interviews we learned that the student administrator and study counsellor at CSE definitely receive a larger amount of email from students for handling requests, compared to employees not related to student affairs. Being able to see the emailing frequency of members of the network would also show which people use email very infrequently, which could be an indicator of **F7 Mismatch of team's styles of communication.** This information could also help imply areas where **E3: Intense communication** and **E5 High interaction frequency** are present. When calculating the speed of these interactions, **F6 Lack of common time** should be taken into account in order to get more realistic results, since when people respond to email is affected by timezones, weekends and holidays. This factor may not affect email networks as much as the others however, due to the asynchronous nature of email (for example how recipients do not have to stop what they are doing when you send a message, unlike answering a phonecall or going to a meeting). Interviewees have also emphasized how if something needs to be

handled/discussed straight away, other communication methods such as a phonecall may be preferable to email. The visualization could easily provide the information necessary to communicate with people in other ways, (such as phone numbers) as long as consent has been given. If the system were to provide **R6 Shared calendars** it should also be possible to reduce the risk of **E1 Waiting for communication** and **E2 Waiting for value**, since people could use these calendars to see when a certain person most likely won't be able to answer emails (knowing this may lead them to look for someone else to email instead). In regards to **F5 No Co-location**, This issue could be improved by having the visualization provide the necessary information for physically finding a certain person if needed (unless the member in the organization that you are looking for is on an offshore location).

In order to improve the awareness of important email for both senders and receivers (related to **R7 Creating awareness**), a part of the visualization could also take **F9 Low prioritized interaction** into account. For example, being able to see if one has unhandled high priority email through the visualization would be convenient. With a ranking system for priority, the recipient should also be made aware of the importance of the email, which may lead to it being handled faster. However, this kind of functionality may have too much overlap with the actual email clients used by people in the organization, and could also cause issues in regards to privacy.

APPENDIX E: INTERVIEW QUESTIONS

Open-ended questions used in semi-structured interviews

1. Can you describe how you communicate with colleagues/students at CSE?

2. What would you consider to be the good points of communication at CSE?

3. Is there anything that you think could use improvements?

4. How frequently do you use email to communicate (in CSE related matters)?

5. What would you consider to be a fast email communication?

6. What information would you be interested in knowing about students/colleagues and your communication with them?

7. The pictures below are example representations of the most common way to model a communication network (graph diagrams where the nodes represent people and the edges represent their relations). Given such a representation, how would you like to interact with or manipulate it?
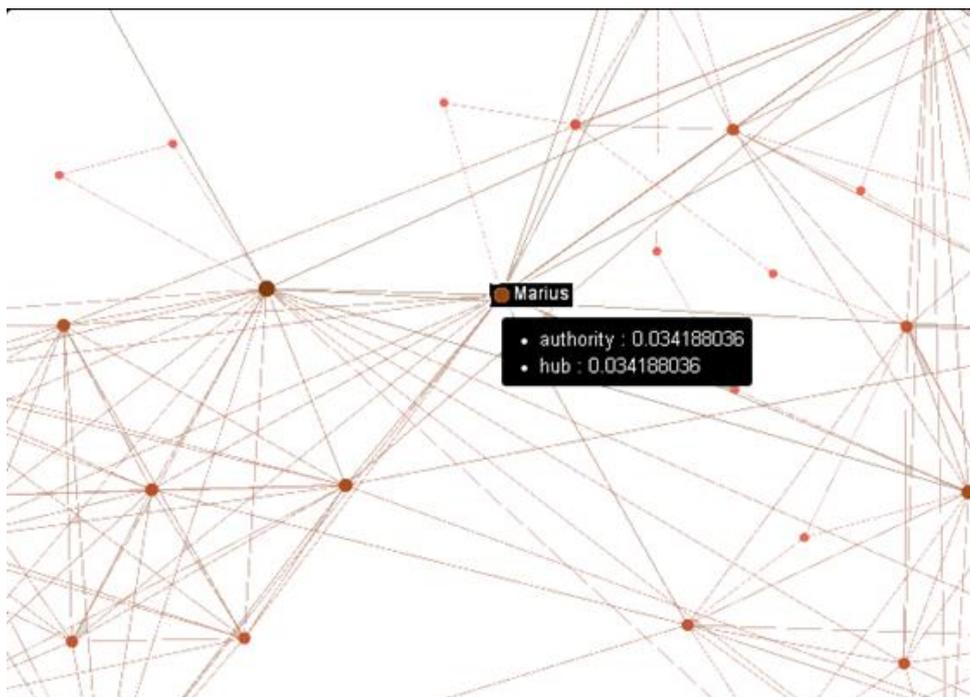
Questionnaire used during iteration 2 structured interviews

1. Basic visualization

1.1 We decided that the visualization of email social networks should be web-based, due to how it allows for easy distribution and access for anyone who wishes to view this information. Would you agree with this choice?

A) Yes.
B) Possibly, but...
C) No, another platform should be used.

1.2 We use different colours on both nodes and edges to represent certain attributes (edge colour = average reply speed, node colour = role). Would you agree with how we currently use colours?

A) Yes it is acceptable.
B) No, colours should be used differently.
C) No, colours should be used less.

1.3 We also change the size of nodes and the width of edges (width of edge = quantity of email sent to target, size of node = amount of email person has sent & received). Would you agree with how we currently use size and shape?

A) Yes it is acceptable.
B) No, size and shape should be used differently.
C) No, size and shape should be used less.

1.4 Currently, we plan to use two edges for every connection, in order to individually visualize interaction speed statistics. However, this can make the visualization a bit messy since twice the amount of edges is needed. Do you think our solution for showing the interaction speed for both sides in a relationship is viable?

A) Yes it is acceptable.
B) It is sufficient but needs some minor changes.
C) Another solution should be found.

2. Interaction statistics

2.1 We are measuring the amount of email, level of activity and the reply speed for each direction in each connection. Would you agree that this is enough to determine how these two people communicate via email?

A) Yes, it is enough.
B) Possibly, but unsure.
C) No, something else needs to be taken into consideration.

2.2 The statistics we show for each person's "contribution" to a single conversation is their percentage of the total amount of messages, if they started the conversation and their average reply speed in this conversation. Would you agree that this is enough to measure that person's participation in a single conversation?

A) Yes, it is enough.
B) Possibly, but unsure.
C) No, something else needs to be taken into consideration.

2.3 We are considering any reply speed that is below 24 hours as fast (due to the asynchronous nature of email). Would you agree with this time limit?

A) Yes, it is acceptable.
B) Possibly, but unsure.
C) No, it is too high/low.

3. User page
3.1 We plan to visualize personal information such as role and location on campus. Is there anything else that you think would be good to know about this person?

A) No, not really.
B) Perhaps.
C) Yes, something else should also be shown.

3.2 We are considering having "free-text" sections on each user's page. The purpose of this is to allow users to show more information about themselves. Would you agree that this is useful?

A) Yes, it is useful.
B) Possibly, but unsure.
C) No, additional information is not necessary.

3.3 We also plan on displaying calendars on each user's page. The purpose of this is to allow users to show when they are available/busy. Would this is useful?

A) Yes, it is useful.
B) Possibly, but unsure.
C) No, creating these calendars is too much work.

4. Tag system
4.1 Nodes show more information when hovering over them. This information includes name, role and tags of the person represented by the node. Is the information enough for giving basic knowledge of the person in question?

A) Yes, It is sufficient.
B) No, there's too little information.
C) No, there's too much information.

4.2 How do you like the arrangement of node information in the panel?

A) Good.
B) Needs slight changes.
C) Bad.

4.3 We plan to allow users to add searchable tags to both nodes and edges, so that they can be used for filtering the visualization. Would you agree that this kind of tag feature is helpful, or does it rely too much on user contribution?

A) Yes, it is acceptable.
B) Possibly, but unsure.
C) No, the users should not have to manually add information to the visualization.

4.4 How do you like the way that tags are added and represented in the mockup?

A) Good.
B) Needs slight changes.
C) Bad.

5. Email Conversation Visualization
5.1 We plan to differentiate between handled/unhandled email conversations via icons (and showing the people (nodes) connected to these conversations). How do you feel about using different icons to represent the status of emails?

A) Yes.
B) Possibly.
C) No, it is unrelated.

5.2 We also want to visualize the priority of email conversations. our current idea for it would require users to insert certain keywords in their email subjects to represent a certain level of priority (or add a priority tag to the conversation via our visualization). Do you think such a priority system would be feasible?

A) Yes, it is feasible.
B) Possibly, but unsure.
C) No, it relies too much on user action.

5.3 What do think of using colour to represent certain levels of priority?

A) Good.
B) Needs slight changes.
C) Bad.

5.4 Are two levels of priority enough?

A) Yes it is enough.
B) No, two is too few.
C) Priority system should not be used (since it may be abused).

6. Filtering

6.1 Currently, we plan to support filtering via searchable node/edge tags, node relations, time spans, and custom groups. Would you agree that we are providing enough ways to filter the visualization?

A) Yes, it is enough.
B) No, there are too few filters.
C) No, there are too many filters.

6.2 The effects of filtering (besides timespan filtering) are displayed by having "unrelated" elements lose their colour. Is this a good way of showing the effects of the currently used filter(s)?

A) Yes, it is acceptable.
B) Possibly, but unsure.
C) No, the "unrelated" elements should not be visible at all.