# MITeams:

## Quick Organizational Mapping by Combining Email and Survey Data

**Jingxian Zhang**

M.S. in Computer Science
University of Illinois at Urbana-Champaign, 2016

Bachelor in Engineering
University of Science and Technology of China, 2014

Submitted to the
**Program in Media Arts and Sciences,**
**School of Architecture and Planning,**

in Partial Fulfillment of the Requirements for the Degree of

**Master of Science in Media Arts and Sciences**
at the **Massachusetts Institute of Technology**

June 2018


**Author** Signature redacted

Jingxian Zhang
Program in Media Arts and Sciences
May 4, 2018

**Certified By** Signature redacted

Prof. Cesar A. Hidalgo
Associate Professor in Media Arts and Sciences
Program in Media Arts and Sciences

**Accepted By** Signature redacted

Prof. Tod Machover
Academic Head
Program in Media Arts and Sciences

# MITeams:
## Quick Organizational Mapping by Combining Email and Survey Data

**Jingxian Zhang**

Submitted to the
**Program in Media Arts and Sciences,**
**School of Architecture and Planning,**

in Partial Fulfillment of the Requirements for the Degree of

**Master of Science in Media Arts and Sciences**
at the **Massachusetts Institute of Technology**

Signature redacted

**Advisor**

**Prof. Cesar A. Hidalgo**
Associate Professor in Media Arts and Sciences
Program in Media Arts and Sciences

Signature redacted

**Reader**

**Prof. Alex "Sandy" Pentland**
Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Signature redacted

**Reader**

**Prof. Kent Larson**
Principal Research Scientist of Media Arts and Sciences
Program in Media Arts and Sciences

# Acknowledgement

I would like to thank my advisor, Cesar Hidalgo, for his consistent support and intellectual guidance during the two years. Without his support this project would not have been possible. He is one of the smartest people I know and I feel extremely lucky to have worked with him.

Thank my readers, Sandy Pentland and Kent Larson, for their support and guidance during the process. Their knowledge and insights inspire me and bring my thesis project to a new level.

Huge thanks to my colleagues in Collective Learning and the Media Lab, for I can always get quick and warm advice from them. It's their support and passion that make my two years in the Media Lab a joyful and colorful experience. I love you all.

Finally I would like to thank my family and friends for being there when I need them and being supportive to nearly all my decisions.

# Contents

# MITeams: Quick Organizational Mapping by Combining Email and Survey Data

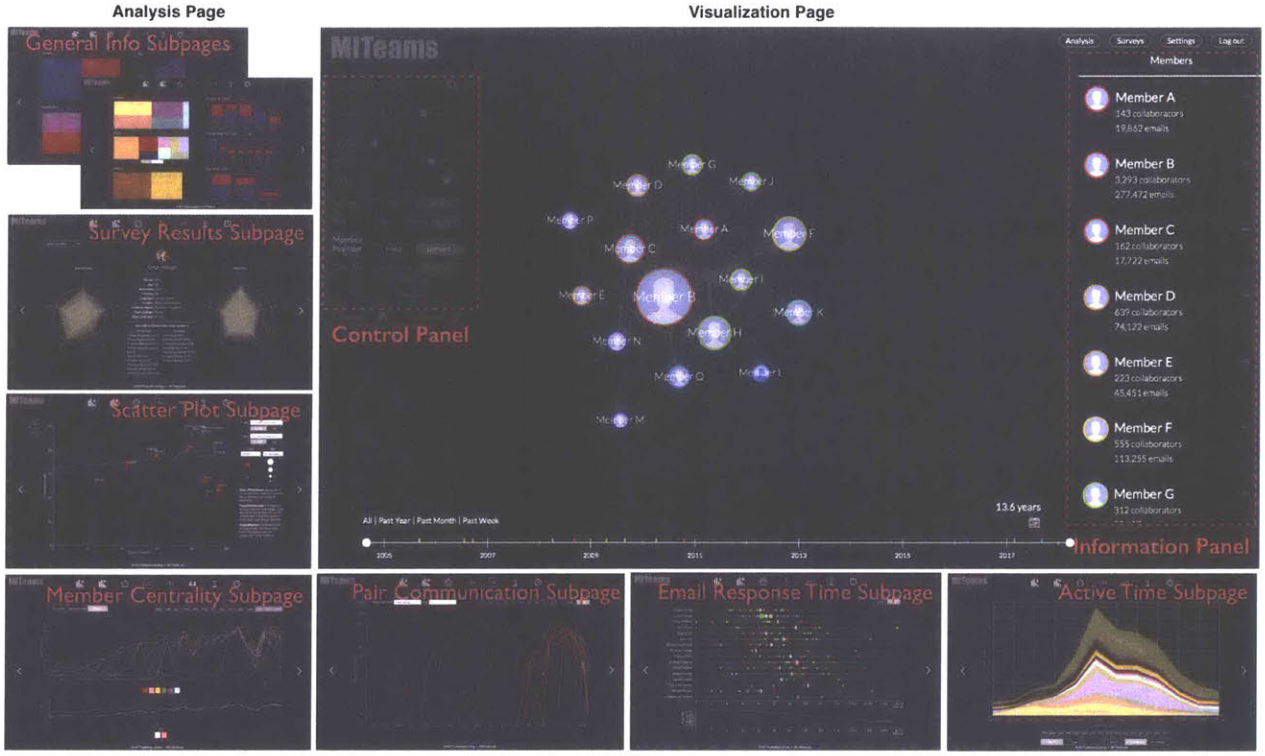



**Figure 1: The MITeams interfaces. The Visualization page consists of a control panel on the left, an information panel on the right, a main visualization in the center, and the timeline slider on the bottom. The Analysis page contains eight subpages: two general info subpages containing demographic and work info, a survey results subpage with members' survey results in radar charts, a scatter plot subpage presenting the distribution of selected attributes of members (with an option to download data), a member centrality subpage presenting the change of member centrality and team density over time, a pair communication subpage displaying communication strength between pairs of members by time, an email response time subpage presenting how team members reply to and get replied by their colleagues, and at last, an active time subpage showing the active hours and days of members.**


## ABSTRACT

Organizational maps can help teams and organizations improve how they manage their human and social capital. Yet, keeping track of the relationships in an organization is a complex, labor-intensive, and time-consuming task. Here I present MITeams: an automated visualization tool for quick organizational mapping. MITeams uses email metadata to map the pattern of communications within an team and combines this information with demographic and psychological variables collected from externally validated surveys included in MITeams. These capacities allow MITeams to create maps of teams and organizations, combining communication and psychological variables, in a matter of hours. I present several use cases of MITeams with the user experience of a research team. User study is conducted with another research team to validate MITeams's the ability in mapping the communication pattern and assisting members learn team knowledge. Our results show that MITeams can be used to quickly map organizations, and help their members get a deeper understanding about team communication network and dynamics in a short time.

## 1 INTRODUCTION

Understanding the communication network in teams or organizations can help organizations improve the management of their human and social capital. However, keeping track of the relationships in an organization is labor-intensive and time-consuming. Researching the structure of communication and relationship in a network requires conducting interviews and long surveys, followed by slow analyses and report preparation. This process becomes even more complex when the organizations are larger. Even though the importance of organizational mapping is acknowledged in the human resources and organizational psychology fields, the processes of organizational mapping are mostly inefficient.

For a long time, social network analysis has been used to analyze the relationships in organizations, usually trough conventional methods such as interviews and surveys. A common approach to collect data for social network analysis is by generating pairs of names and for each dyad of actors (nodes) asking a list of questions in order to access the strength of the link between the two. Even though these data can be collected digitally, because it is done at the dyad level, it can become a very long process depending on the number of actors [1]. As one of the main means of communications in organizations, email can be a promising resource for extracting hidden patterns of collaborations in informal networks [2, 3]. Comparing to collecting relationship network data from interviews and surveys, email data has the advantage of being quick and easy to access. They have been used in studying the structure of organizations and are proved to reflect actual social relationships [4, 5, 6, 7].

When it comes to the individuals composing organizations, psychological research suggests that individual variables are usually related with community roles and centrality of individuals in social networks [8]. Personality and moral foundations, in particular, as probably the most relevant psychological constructs and as the engines that drives our actions, are important factors to consider. Thus, putting the information obtained via validated personality and moral foundation inventories together with the information obtained via informal social network can be valuable in understanding the relationships in teams and organizations.

In this thesis, I present MITeams, an automated visualization tool for quick organizational mapping. MITeams uses email metadata to map the pattern of communications within an organization and combines this information with demographic and psychological variables (externally validated). With a Visualization page and an Analysis page, the tool enables users to explore the contact/organization networks as well as team dynamics. I present use cases of MITeams with the user experience of a research team of 16 people. To validate MITeams, a user study is done using MITeams to map the collaboration history and psychological profiles for a research team of 10 people. The user study results demonstrates that MITeams can be used to quickly map organizations, by providing introspection on personal and team communications and helping team members learn team knowledge.

The rest part of this paper is organized as follows. Section 2 describes some related work. Section 3 introduces the data processing in MITeams. The visualization and interactions design of MITeams are summarized in Section 4. Section 5 introduces use cases of MITeams. The conducted user study is presented in detail in Section 6 while the discussion in Section 7. Concluding remarks are presented in Section 8.

## 2 RELATED WORK

Some previous research in social network analysis field used emails and email visualizations. Psychological variables, especially personality, have also been proved useful in the study of team networks. So, in the following section, we present a short review of some of this past work. We also review some approaches and tools previously developed by others for organizational mapping purposes.

### Emails in Social Network Analysis

As a widely popular resource, emails have been used to create team networks, identify communities, examine group interactions, and understand social networks [9, 10]. Empirical analysis with email data have been done studying how the network topology and the organizational structure influence evolving social networks [11, 12]. Social Network Fragments (SNF) and PostHistory are two examples of tools based on email archives [13]. SNF focuses on groups or communities of people that emerge within a person's social network, whereas PostHistory focuses on the social world of dyadic email relationships. SNF uses CC and TO fields to derive a matrix of connections between all the recipients, whose position, is then decided, on a 2D plane, based on a spring system algorithm. PostHistory is a visualization that represents time in a calendar-like format and contains a timeline where more intense and less intense moments of message exchange between the individual and each one of their contacts can be observed [14]. Similarly, Immersion is another tool that was created as a personal email visualization platform that provides a person-centered view of the email network [15, 16]. When compared with SNF, PostHistory, and Immersion, MITeams contains a combination of their main features, the community detection and the timeline. More importantly, instead of focusing on the email metadata of one individual, MITeams aggregates the metadata of multiple people in teams or organizations, allowing visualizations of shared networks. As a quick organizational mapping tool for teams and organizations, MITeams goes beyond the personal email network, aiming to provide insights about team collaboration and team dynamics.

### Psychological Variables and Team Network

Psychological variables, such as personality, are proved to be related with community roles and centrality of individuals in social networks [8]. Personality consists of a dispositional, consistent and long-lasting construct that is responsible for the individual attributes, frequent thoughts, feeling and behaviors [17]. The field of organizational psychology has been studying the impact of personality on job-related variables for a long time and, consequently, a large amount of personality inventories have been developed (Myers-Briggs Type Indicator, the Minnesota Multiphasic Personality Inventory, Revised NEO Personality Inventory, and Hogan Personality Inventory are

some examples). Many of the inventories are based on the Big Five Model of personality, where it is claimed that the personality traits tend to cluster in 5 dimensions: Neuroticism (sensitive/nervous vs. secure/confident); Extraversion (outgoing/energetic vs. solitary/reserved); Open-mindedness (inventive/curious vs. consistent/cautious); Agreeableness (friendly/compassionate vs. challenging/detached); and Conscientiousness: (efficient/organized vs. easy-going/careless) [18]. The personalities are believed to effect individual behaviors in organizations, team effectiveness (*i.e.*, performance) and viability (*i.e.*, and capability to continue working together), and team's network structure [19, 20, 21, 22]. Based on the past research, we believe that by combining psychological variables with team collaboration network, a more comprehensive organizational mapping approach can be developed. As such, MITeams is the results of an effort to combine these features in order to better understand team structure and team dynamics.

### Organizational Network Mapping

Researchers have been trying to understand the relationships in organizational networks using various methods and sources. Surveys and interviews have been widely used for obtaining social ties in organizational networks [23]. Organizational mapping approaches such as Renga and Net-Map, are designed for collecting organizational data using workshops and interviews [24, 25]. In traditional survey methods, data collection and data analysis are usually slow and labor-intensive, with a time gap between the two steps. Compared to the above methods, MITeams would improve the experience by providing automated data collection and visualization, which minimize the time gap between steps.

Apart from collecting data from employees with survey and interviews, emails and other online social network sources have also been studied in understanding the network of relationships in organization [10]. Researchers studied the strategies for selecting mailboxes and found a good approximation of group network could be obtained by using only one fourth of the mailboxes [26]. By studying email and other sources, Gloor identified several indicators of team collaboration, such as type of leadership and response time, that can guide managers to organize and lead teams [27]. Other factors such as co-location can influence people's preference for using face-to-face interactions or emails [28]. These work provides insights for MITeams on what attributes to collect and analyze.

As an important form of social interaction in organizations, face-to-face interaction has been collected and analyzed with different methods. Wearable badges such as the Active Badge, Thinking Tags, iBadge, and the Sociometric Badge, have been developing from recording location and simple motions to complex patterns of behaviours [29, 30, 31, 32, 33]. MITeams doesn't explore the face-to-face channel in social interaction, but could be served as a complimentary platform to such wearable badges, achieving an automated collection of both physical and electronic interaction in an organization.

## 3 DATA PROCESSING

In this section, I'll introduce what data are collected in MITeams, how networks are constructed from email headers, and what metrics are used for network analysis in MITeams.

### Data Collection

The data used in MITeams include email headers, which are automatically collected when users log in using their email accounts, and survey data that are collected when users answers the intergated surveys in MITeams.

#### *Email Metadata*

Teams can use MITeams by creating rooms and have their members join the rooms. To join a room in MITeams, users are required to login through an email account. There are three login options users can choose from – Gmail, Microsoft Outlook, and Microsoft Exchange. By agreeing to login, users give MITeams the permission to access the user profile and email header data. Email content and subject are not accessed in the process. The FROM, TO, CC fields, and the timestamp of emails are extracted from email headers for the construction of contact and organization networks. MITeams users have the option to delete their data when logging out of the platform. When this option is selected, users' profile and email data will be erased from the database, as they will automatically exit all the MITeams rooms they have joined.

#### *Survey Data*

To provide team members with a deeper understanding of team composition, there are three surveys integrated in MITeams: a demographic survey, a personality survey, and a morality survey [34, 35]. For demographic data, team members are asked to provide age, gender, nationality, ethnicity, majors, academic degree, and office location. MITeams provides the possibility of taking a personality inventory, based on the Big Five Model. In this inventory participants are evaluated on five personality dimensions - Open-mindedness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Team members can get their personality scores right after answering 60 questions in the survey [34]. The moral foundations survey with 30 questions measures people's reliance on and endorsement of five psychological foundations of morality: Fairness, Purity, Authority, Harm, and Loyalty [35]. The results of the personality and morality survey can be explored in the Analysis page under the survey results subpage. It is also used as an option to color member nodes in the network.

### Constructing networks

MITeams constructs a *contact network* and an *organization network* using email headers. Each email header contains FROM, TO, and CC fields, which corresponds to the email sender and recipients, and timestamp of the email. For each team member, by going through all their email headers, a list of email contacts will be generated along with the number of emails exchanges between the team member and their every contact. This number can be seen as a presentation for the communication strength. Contacts may use various email addresses during different times, so getting correct and complete email aliases for contacts is essential for merging contacts and creating an accurate combined network. By matching the

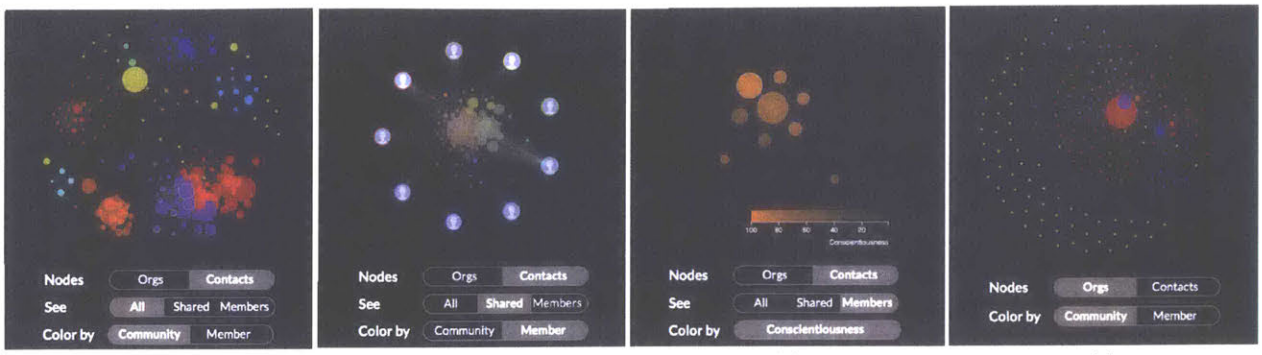



**Figure 2: The contact and organization networks with different coloring and node filtering methods. (a) Contact network showing all nodes colored by community detection results, with member node positions unfixed. (b) Contact network showing shared contacts colored by member, with member nodes in fixed positions. (c) Contact network only showing members colored by conscientiousness. (d) Organization network colored by community detection results.**

names, a list of email addresses are obtained for each contact. When a contact is shared by more than one team member, the data about the contact from all these members will be merged.

In the contact network, each person (a contact or a team member) is a node. The communication strength between two people is the weight of the link connecting the two nodes. Here, the communication strength between $node_i$ and $node_j$ is calculated as the power mean with exponent 3 of emails $node_i$ has sent to $node_j$ and $node_j$ has sent to $node_i$ [15, 16], using equation 1.

$$linkWeight_{ij} = \left[\frac{email_{ij}^{-3} + email_{ji}^{-3}}{2}\right]^{-\frac{1}{3}} \quad (1)$$

Similarly, the weight of each node is the generalized mean of the number of emails the contact has sent to and received from all the team members (equation 2).

$$nodeWeight_i = \left[\frac{sent_i^{-3} + received_i^{-3}}{2}\right]^{-\frac{1}{3}} \quad (2)$$

Contacts that are shared by more than one team member usually have, on average, higher weights than contacts only linked to one team member.

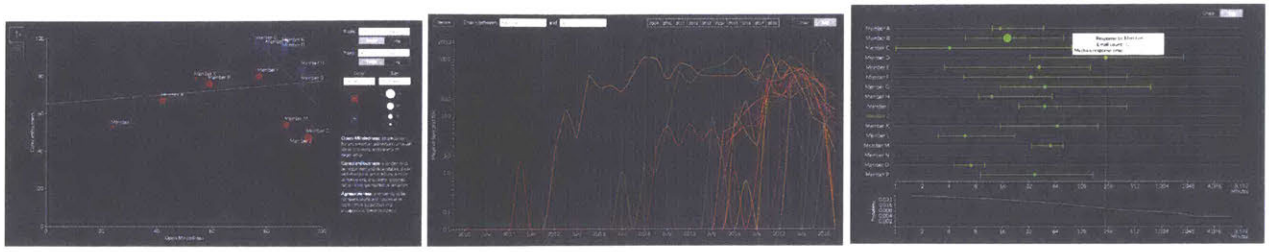
Email domains contain information about which organization this address is from. After pruning email addresses that are from social medias and other online services, a list of organizations is assigned to each contact based on the domain in his/her email aliases. Next, an organization network is constructed, with nodes as organizations and links as the communication strength between people in different organizations. For instance, contact $C_1$ sent an email to contact $C_2$ using an email address from organization $O_1$. This email was sent to $C_2$'s email address from organization $O_2$. Then, the communication volume of $O_1$ to $O_2$ would be increased by 1. The link and node weights in the organization network follow equation 1 and 2. The organization network represents how people in the team communicate from an organizational view.

In order to see which contacts or organizations are closely connected internally, a modularity-based community detection algorithm is applied to both the contact network and the organization network to divide contacts and organizations into communities [36]. The community detection results can be presented by changing the coloring method to *by community* (see Figure 2).

### Data Analysis

As important measures of the influence of nodes within a network, three types of centrality are used in MITeams to present the importance of nodes and also how the importance changes over time. Normalized degree, closeness, and betweenness centrality are computed based on a 60-day time window for nodes in the contact network, the results of which are presented in the member centrality subpage.The centrality ranking of contacts and organizations in the Visualization page is based on betweenness centrality. When the team composition changes, the changes in the ranking are marked in red and green arrows (see Figure 4). Densities of the contact network and the largest component in it based on a 60-day time window are also computed and presented in the Analysis page.

For teams that participated in the user study, an analysis is done to see the relationship between the link weight (communication strength) and homophily of nodes. In the analysis, contact networks are constructed based on the email communication within a window 60 days. Dimensions from demographic (gender, age, academic degree, and job position), personality survey, and morality survey are used as attributes for nodes. The analysis is based on stochastic actor-based models for network dynamics where an R package SIENA is used for it [37, 38]. Results of the analysis for teams can be found in the Discussion section. As part of the homophily, similarity of members based on personality and moral foundations are demonstrated in the survey results subpage.



(a) (b) (c)

**Figure 3: The scatter plot subpage, pair communication subpage, and response time subpage in the Analysis page. (a) shows a scatter plot with the x-axis selected as Open-mindedness, the y-axis as Conscientiousness, the color as Gender, and the size as Agreeableness. (b) shows how the email communication between Member *B* and other members change over time. (c) is the response time visualization highlighting how people in the team reply to Member *J*.**

To understand team network, paying attention to how the communication between members changes over time is crucial. For this reason, MITeams computes the communication volume (presented by the arithmetic mean) of emails exchanged in every three months for every pair of members that are connected in the network. A line chart of the communication volume change is presented in the pair communication subpage in the Analysis page. In the Visualization page, MITeams creates, for each member, a ranking of their contacts based on the communication volume.

Knowing the active time of members in your team can be helpful in learning the best time to reach a colleague via email. Using timestamps in email headers, the active time of users sending emails is calculated in MITeams. The distribution of users' active hours in a day (based on 2-hour time slots) and active days in a week (Monday to Sunday) are presented in the active time subpage under the Analysis page.

The email response time to members is another important feature in understanding the communication pattern in a team. After sorting emails with the thread ID and finding matches of names in FROM, TO, and CC fields, I get the response time between member pairs in a team, the results of which is shown in the response time subpage under the Analysis page.

### Privacy Related

In MITeams, only email headers are used to construct networks. Email content and subject are not accessed in the platform. Some user data is saved in the database when a user is visiting MITeams, which they can always delete when they log out. When users log out of a room in MITeams, they are given three options: (1) log out and save the data for next use, (2) exit the room but save the data for other rooms he/she has joined, and (3) exit all rooms and delete all his/her data. If option 1 is selected, the user's data will be kept in the MITeams database for his/her next visit. If the user wants to be erased from the room but keep his/her data in the database because he/she has joined other rooms, the user can choose option 2. To eliminate his/her name from all rooms and delete his/her data from the database, user could choose the third option.

## 4 VISUALIZATION AND INTERACTION DESIGN

The MITeams interface consists of two main pages – a *Visualization page* and an Analysis page. After logging in, team members will be directed to the Visualization page which contains the network visualization, information panel, and the control panel where they can tune the network and visualization settings. More analysis of the team network can be obtained in the Analysis page, which can be assessed from the analysis button on the Visualization page (see Figure 1).

### Implementation

MITeams is a web application built using Javascript, HTML, and CSS in the frontend and Python in the backend. The frontend is responsible for constructing networks, visualization, and part of data analysis, while the backend takes care of login authentication, email fetching, email parsing, data cleaning, data storage, and most of data analysis [15, 16]. Data in MITeams are stored in MongoDB and the visualizations are created using D3.js and D3plus.js [39, 40].

### Visualization Page

The Visualization page consists of three parts: a network visualization in the center, a control panel on the left, and an information panel on the right (Figure 1). On the control panel, users can tune the basic network settings, such as number of nodes/links being displayed and the minimal distance between nodes, using the three sliders on top. For instance, if users want to see contacts with high communication volume, they can use the node slider to filter nodes by weight. Below the sliders are three fields with buttons, one that enable users to switch between the contact and the organization network, one that filters the nodes by contacts, shared contact, and team members, and one that colors the nodes based on six dimensions: members they are connected to, the community they belong to, and selected dimensions in the personality inventory (Figure 2). The search box on the top of the control panel is for contact and organization search in the network, where the search results can be highlighted.

Below the network visualization, the timeline slider enables users to change the time period and reconstruct the network correspondingly. Options such as past week, past month, and

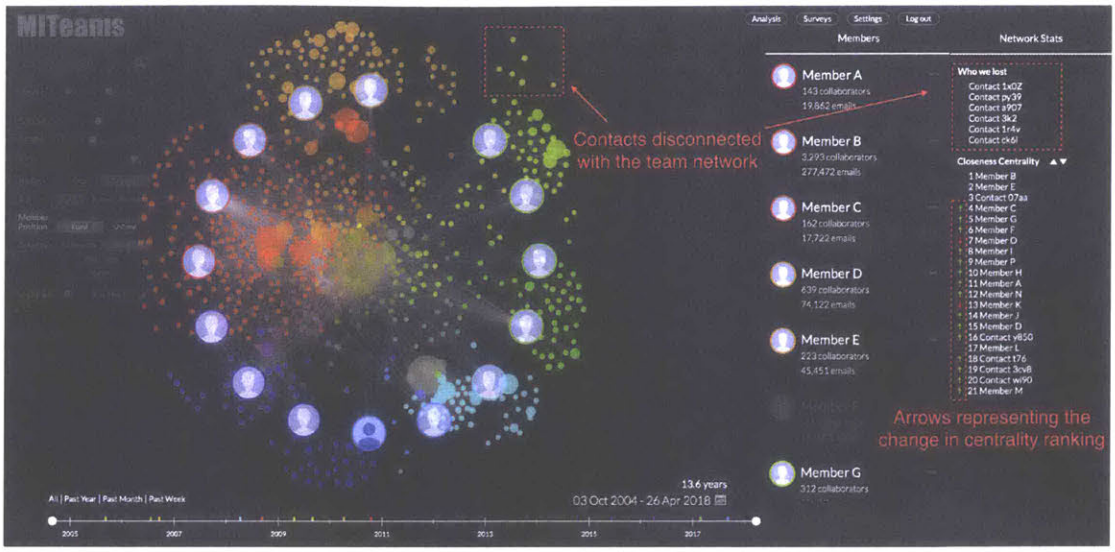



**Figure 4: The team network when member $E$ is removed. Nodes in the red dashed box on top are the contacts that are disconnected with the main network after member $E$ is removed. The centrality ranking on the information panel on right changes with the member's departure. Red arrows represent ranking going down and green ones represent ranking going up.**

past year are listed for quickly selecting a time period. As shown in the Visualization page in Figure 1, a list of team members is displayed on the information panel on the right. Each member has a name card with their name, number of collaborators, and number of emails on it. Contacts of a member will be highlighted when their name card is hovered over. When a name card is clicked, statistics about emails the member sent, received, and the number of new collaborators per year will be presented in the information panel. Betweenness centrality of team members are ranked in the right part in the information panel. The minus button on a name card allows users to "remove" a member from the team and see changes in contacts and centrality ranking from the information panel (Figure 4).

### Analysis Page

By clicking the Analysis button on the top right corner of the Visualization page, users will be directed to the *Analysis page*. Analysis page contains more detailed information about the communication and dynamics within the team. The page consists of eight subpages, namely the demographic info subpage, work info subpage, survey results subpage, member centrality subpage, scatter plot subpage, pair communication subpage, response time subpage, and active time subpage. For most subpages, time sliders and buttons to switch between linear and log scale are implemented for the best experience of exploring the visualization.

*Demographic info subpage* contains tree maps and histograms presenting the distribution of gender, age, and nationality of team members.

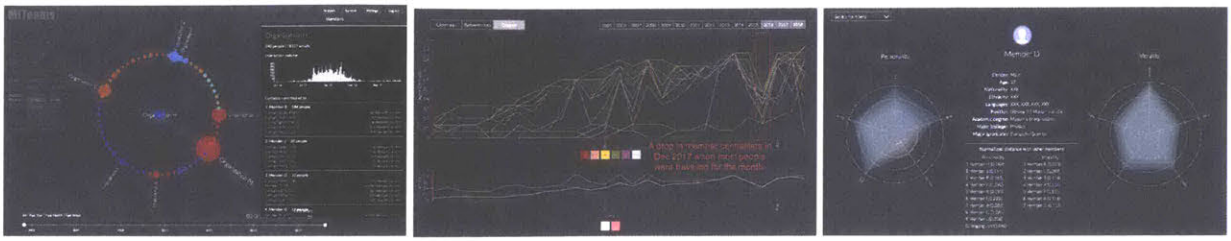
*Work info subpage* contains tree maps and histograms demonstrating the distribution of members' work related dimensions: the job position, major, and academic degree. In the histograms, users could select a second dimension and see how position, major, or academic degree in the team distributes by the second dimension.

*Survey results subpage* uses two radar chart visualizing the personality and morality survey results of the team. For easy comparison between members, a drop down list containing member names is implemented for member selection. When a single member is selected, his/her similarity (Euclidean distance) with the other members in both survey results will be presented in the page. An example of survey results subpage is presented in Figure 5c.

*Scatter plot subpage* displays members as dots in a 2D coordinate space. It presents users how members are distributed in two selected dimensions. The color and size of a dot also represent dimensions selected by users (see Figure 3a). When gender is selected for color, trendlines will be plotted demonstrating the correlation between selected dimensions for female, male, and all genders.

*Member Centrality subpage* uses line charts to present how centrality of members (closeness, betweenness, and degree) and density of the team network change over time. The lines are colored by the job position of members, which aims to give users insights on how people from different positions are connected with others in a team. An example of member centrality subpage is presented in Figure 5b.

*Pair communication subpage*. In the pair communication subpage, each line represents the communication between a



(a) (b) (c)

**Figure 5: The organization network in the Visualization page, member centrality subpage, and survey results subpage in the Analysis page. (a) shows members who have contacts in Organization $H$ and how it's connected with other organizations. (b) shows how member centrality changed over time where a drop could be noticed in Dec 2017 when most people in the group were traveling. (c) presents the similarity in personality and moral foundations between members.**

pair of members that are connected in the team in every three months. Users can filter out the communication line for one pair or lines for all pairs related to a member in the dropdown menu (see Figure 3b). The table subpage presents the above data of members in a table format, which can be downloaded in a CSV file for further use.

*Response time subpage* uses a modified box plot demonstrating the email response time of every member in the team (Figure 3c). In this subpage, each member is represented by a color. For each row in the chart, the colored circles and bars presents how the member replies to the other members in the team. The rhombus in each row presents the median response time of the member.

*Active time subpage* uses stacked layer chart presenting members' active hours in a day (2-hour time slots) and active days (Monday to Sunday) in a week. The Y axis in the visualization can be chosen from number of emails, number of emails per day, and percentage of emails.

## 5 USE CASES

A research group of 16 people used MITeams in group meetings and individually multiple times. The research group consists of one PI, one development director, six graduate students (four master's and two PhD students), four postdocs, three visiting PhD students, and one visiting professor. In this section, this team is used as an example to demonstrate some use cases of MITeams. The names of members and contacts are coded with the consideration of their privacy.

*When a member leaves the team.* Teams may encounter situations in which a member leaves the team. It is usually difficult to evaluate or predict how such a situation can impact the team network. In MITeams, users can simulate this situation and see how the network will change after one member leaves, how are the connections between the remaining members affected, what are the external social capital that is being lost. Member $F$ is a senior graduate student in the team who is about to graduate. By clicking on the minus button on his name card in the information panel, we can "remove" member $F$ from the team. As shown in Figure 4, after member $F$ is removed from the network, some of their contacts are disconnected and floating far from the team network. These contacts, if valuable for the team, are better to be reached and connected by other team members before member $F$ leaves the team. The information panel on right presents the betweenness centrality ranking of nodes after member $F$ is removed, with red up and green down arrows showing how the centrality of a node has been changed by member $F$'s departure.

*Reach out to a new organization.* Member $A$ is a master's student in the group working on a project to promote team collaboration. She would like to make contact with experts in Organization $H$ working on the same topic to gain more related knowledge. In order to see which colleague in the team could be a good referrer, Member $A$ clicks the node Organization $H$ in the organization network. Details about which members have contacts in the organization are listed in the information panel, with which Member $A$ knows from which colleague she should ask to help connect (Figure 5a).

*Introspection on past events.* With the aim to learn when the team is most and least connected, the team explores the network centrality subpage to see how centrality of members and density of the whole team network change over time. It comes to their attention that there is a drop in the closeness and degree centrality of many members in Dec 2017 (Figure **??**). Recalling what happened that time, it turned out that is a month when the PI of the group was traveling and most students planned long vacations. From the centrality visualization, there is a purple line whose centrality increased fast in September 2017 and keeps on the top since then. This member is the director of development in the group who joined the team in September 2017. After her joining, she keeps close communication with every member in the group, which is reflected from the plot.

*Learn more about your colleagues.* After finishing the demographic, personality, and morality surveys, Member $C$ explores the Analysis Page to learn more about the background of him team. He sees what majors his colleagues are from, which gives him insights about who to consult in the future for questions in different fields of study. From the survey results subpage, he sees who are more similar to him in personality dimensions (Figure 5c). In the active time subpage, Member $C$ learns that the most active hours for the PI is 10-12am and

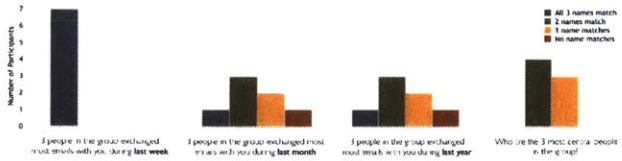



**Figure 6: The comparison of answers to a set of same questions in the pre-study survey and the main survey. Participants have clear memory about their communication not long ago (last week) but may need help to recall accurate communication farther than that.**

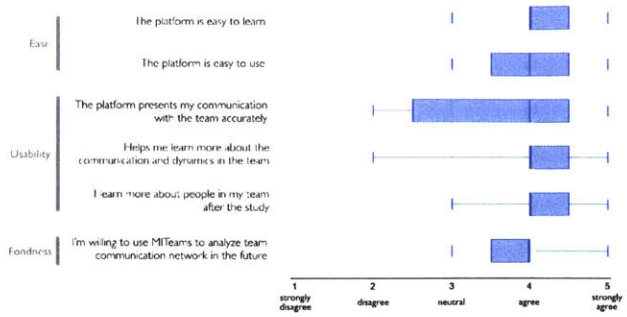



**Figure 7: Results of the post-study questionnaire about the ease of use, usability, and fondness of MITeams.**

the next is 2-4pm, which gives him an idea of when could be the best time to contact the PI.

## 6 USER STUDY

Apart from the group in the Use Cases Section, user study is conducted with another research group evaluate the usability of MITeams. In the study, participants from a group are required to first answer a pre-study survey containing questions collecting their perception of the group; then they'll watch a demo demonstrating the interfaces and user interactions in MITeams; after the demo, they will explore their team network in MITeams while answering survey questions about their team communication; finally, questionnaires are performed to collect the participants' feedback about the experience of using the platform. In the user study, apart from participants who are in the study, all contacts info (email addresses and names) are made anonymous. They show up in the platform with coded hashing ids such as *Contact Id8z0*.

### Participants

The research group consists of 10 people (6 males and 4 females), including 1 PI, 1 administrative staff, 2 PhD students, and 6 master's students from six countries.

### Tasks

The user study consists of 4 parts: the pre-study survey, demo, main survey, and post-study questionnaire. Questions in the surveys are listed in Appendix I. The pre-study survey collects participants' perception about the group before using MITeams. It contains questions like *which 3 people in the group do you think exchanged most emails with you during last month* and *to which 3 people in the group do you respond the fastest*.

The demo part serves as a tutorial for participants to learn how the platform works. In the demo, participants will learn all the pages and user interactions in MITeams while watching the investigator present the communication network of a sample team (the research team in Use Cases section).

In the main survey part, participants will explore their team network in MITeams while answering the main survey, which contains similar questions in the pre-study survey. These questions aims to collect participants' perception about the team after using MITeams.

The post-study questionnaire collects user feedback regarding their experience using MITeams. It contains 5-point Likert-scale questions as well as open-ended questions from the aspect of ease of use, usability, and fondness.

### Results

In the pre-study survey and the main survey, we have same name-generating questions asking the participants who exchanged most emails with them in last week/month/year, who are more central in the group, and email response time of members in the group. Comparison of participants' answers in both surveys is presented in Figure 6. All the participants give exactly same three names for Q1 in both the pre-study survey and the main survey. When it comes to last month (Q2) and last year (Q3), 4 out of 7 participants give more than two same names. To some extent, it shows people can remember their email communication not long ago. But for those that are far from now, people don't usually have a clear memory and need some help to recall. This is where MITeams could help team members with. Participants are asked to list the nationalities and college majors of members in the team. Before using the tool, 1 participant list all the nationalities of the members correctly, 2 give 83.3% correct nationalities, 2 list 66.7% correct, and the rest 1 gives 50% correct nationalities. For the college majors, 2 participants give 80% correct answers and 4 give 60% correct majors. After using the tool, all participants give 100% correct answers for both questions.

Figure 7 presents results from the post-study questionnaire. Participants feel the platform is easy to learn and use. In the open-ended questions, all the participants reflect that they will be able to use MITeams even without the demo part. On average, participants feel MITeams presents their communication with the team accurately. There are 2 disagree collected for this question because participants think they have more face-to-face interaction with colleagues, which is not presented in the platform, and colleague who have closer communication with them don't participate in the study. It's agreed by most participants that MITeams helps them learn more about team communication, people in the group, and team dynamics. Participants report willingness to use MITeams in the future to analyze team communication network (1 strongly agree, 5 agree, and 2 neutral).

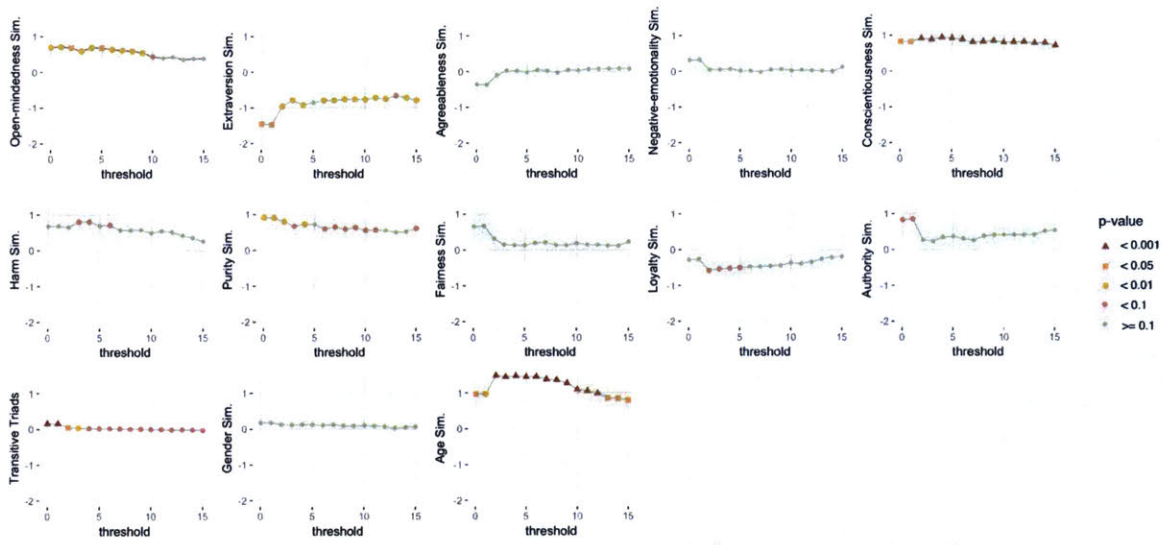



**Figure 8: Parameters and significance levels of different attributes of members in email communication network with the threshold of forming a link varying from 0 to 15 emails. Similar age, conscientiousness, and extraversion value appear to be significant attributes for forming a link between members.**

Participants are asked to list features, visualization, or interactions that are most interesting and most difficult for them to explore in MITeams. 4 out of 8 participants report the comparisons to teammates being the most attractive feature to them. 2 participants reflect the email response time visualization is interesting to explore. For the features that are difficult to use, participants report the long waiting time when the time slider is changed and bugs in the response time subpage when more people joined the room. When asked about the potential scenarios of using MITeams in organizations, participants give answers include it can be used for HR and executive managers to understand the dynamics of the team, to understand what clients you will lose when losing a specific employee, and as a tool in team building workshop.

## 7 DISCUSSION

With the aim to explore how the email communication between team members is related with the homophily of them, I look at age, gender, shared contacts, five personality dimensions, and five morality dimensions of the team from the Use Cases Section (team 1) and the team in the user study (team 2). For both teams, email communication networks are constructed based on a time window of 60 days, for consecutive 18 months from September 2016 to February 2018. A stochastic actor-based model is run for both teams using RSiena [37]. For team 1's networks, the threshold for forming a link is changed from exchanging 1 email to 15 emails in the 60-day time window. Teams 2 has less strong links and limited number of members finishing both personality and morality surveys, so the threshold is set to change from 1 to 3 emails to keep a good convergence in the estimation. Results for team 1 are presented in Figure 8. From the results, we can see that similarity in age, conscientiousness, and extraversion keep being significant features for forming a link between members with the threshold changing. Shared contacts and open-mindedness appears to be important feature when the threshold for forming a link is low. Similar moral foundations and gender don't influence the email communication significantly in team 1. For team 2, similarity in age and agreeableness appear to be the two most significant features for email communication. I plan to extend the analysis with larger teams and more attributes of nodes in the future. In next step, the results can be integrated in MITeams interfaces for users to explore.

From the user study and use cases, MITeams demonstrates its capability to perform quick organizational mapping for teams. We believe the emails collected by MITeams can, to a large extent, reflect the non-verbal communication of the team. The visualization and interactions in MITeams provides users with the ability to introspect on team communication, to learn about team dynamics, and to explore how to improve the team structure.

*Email communication, collected by MITeams, reflects real team communication network.* Email is the main resource MITeams uses to construct the team network. The fact that emails can be accessed quickly and parsed easily make them good sources of data for quick organizational mapping. Whether emails can be a proxy for team interactions is something that needs more validation. However, our user study suggests that that might, in fact, be the case. Many of the hypotheses we tested were formulated based on what we knew regarding the teams' real interactions, and the pattern of results obtained with MITeams confirmed many of those hypothesis. Moreover, the feedback we received from the users also suggests that email communications can successfully reflect team communication network.

*MITeams provides introspection on personal and team communication.* In the examples in Use Cases section, MITeams helped team members to understand, share and discuss their communication with other members and contacts. Features such as the histograms and ranking of email communication in the information page, the highlight of personal network when a member's name card is hovered, and the centering of a member node with their contacts surrounding it, all help visually explore one's network. When it comes to the team perspective, exploring member centrality and network density changes, filtering shared contacts and member nodes, centrality ranking of nodes in the team, and pair communication curves in the Analysis page fulfill the goal of helping the members understand team structure and communication.

*MITeams helps team members learn team knowledge and dynamics.* It is reflected in the user study that participants learn more about their colleagues and team communication after using the tool. With the demographic, personality, and morality survey embedded, MITeams presents members with team knowledge and supports them understanding team dynamics. In the Visualization page, when the member network is colored by a selected dimension in the personality survey, users can get insights of how personality, communication strength, and centrality of team members are correlated (Figure 2c). The scatter plot subpage, survey results subpage, together with pair communication subpage in the Analysis page, also supports exploration of such correlations (Figure 3a and 3b). Normally, such information is not easy to quickly obtain, combine and explore. MITeams improves this by quickly collecting and combining survey data with team communication network.

*MITeams helps team members explore how to improve the team structure.* With the interactions in MITeams, users can not only learn the team structure but also explore how to improve it. The Visualization page allows users see if the team are well connected or more dependent on small groups. Combined with the centrality information and pair communication curves, team managers and team leaders, can see if the current team structure fits what they have in mind and figure out which connections should be enhanced to improve the team structure. Furthermore, the possibility, provided by MITeams, of removing a member from the team, can be used to predict how is the internal dynamic going to be disturbed and the external connections (contacts or organizations) that the team will lose after a certain member leaves. This can help managers to take action and solve the eventual future gaps, by finding, for example, a good substitute for the leaving member.

## 8 CONCLUSION AND FUTURE WORK

In this paper, I present MITeams, an automated visualization tool for quick organizational mapping. By combining email metadata with demographic and psychological variables collected in embeded surveys, MITeams maps the pattern of communications within teams. With a Visualization page and an Analysis page, the tool enables users to explore the contact and the organization network as well as to analyze the communication within the team. We validate MITeams by using it to map the collaboration history and psychological profiles for a research group. The user study results suggests that MITeams can provide introspection on personal and team communication, help team members learn team knowledge and dynamics, and help team leaders explore how to improve the team structure.

The email is only one channel of team communication. Communication data from other channels, such as instant messages and face-to-face communication, can be integrated into MITeams in the future work for a thorough understanding of team communication network. To give team members more solid understanding of team dynamics and characteristics, other psychological variables, such as the well-being survey, can be added to MITeams in the future. To focus more on targeting problems in team structure and give solutions to users, I also plan to add features such as event detection and providing suggestions on how to improve team structure in the tool. We also plan, to further validate the tool though more user studies, with more diverse teams and organizations.

## REFERENCES


1. Bernie Hogan, Juan Antonio Carrasco, and Barry Wellman. Visualizing personal networks: working with participant-aided sociograms. *Field Methods*, 19(2):116–144, 2007.
2. Steve Whittaker and Candace Sidner. Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 276–283. ACM, 1996.
3. Barry Wellman. Computer networks as social networks. *Science*, 293(5537):2031–2034, 2001.
4. Ronald S Burt. Models of network structure. *Annual review of sociology*, 6(1):79–141, 1980.
5. Bernardo A Huberman and Tad Hogg. Communities of practice: Performance and evolution. *Computational & Mathematical Organization Theory*, 1(1):73–92, 1995.
6. Lada A Adamic and Eytan Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.
7. Nicolas Ducheneaut and Leon A Watts. In search of coherence: a review of e-mail research. *Human–Computer Interaction*, 20(1-2):11–48, 2005.
8. Sylvia A Morelli, Desmond C Ong, Rucha Makati, Matthew O Jackson, and Jamil Zaki. Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, page 201702155, 2017.
9. Joshua R Tyler, Dennis M Wilkinson, and Bernardo A Huberman. Email as spectroscopy: Automated discovery of community structure within organizations. In *Communities and technologies*, pages 81–96. Springer, 2003.
10. David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009.

11. Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *science*, 311(5757):88–90, 2006.
12. Anders Mollgaard, Ingo Zettler, Jesper Dammeyer, Mogens H Jensen, Sune Lehmann, and Joachim Mathiesen. Measure of node similarity in multilayer networks. *PloS one*, 11(6):e0157436, 2016.
13. Fernanda B Viégas and Judith Donath. Social network visualization: Can we go beyond the graph. In *Workshop on social networks, CSCW*, volume 4, pages 6–10, 2004.
14. Fernanda B Viégas, Danah Boyd, David H Nguyen, Jeffrey Potter, and Judith Donath. Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.
15. Deepak Jagdish. *IMMERSION: a platform for visualization and temporal analysis of email data*. PhD thesis, Massachusetts Institute of Technology, 2014.
16. Daniel Smilkov. *Understanding email communication patterns*. PhD thesis, Massachusetts Institute of Technology, 2014.
17. Ning Li, Murray R Barrick, Ryan D Zimmerman, and Dan S Chiaburu. Retaining the productive employee: The role of personality. *Academy of Management Annals*, 8(1):347–395, 2014.
18. Paul T Costa and Robert R McCrae. Four ways five factors are basic. *Personality and individual differences*, 13(6):653–665, 1992.
19. In-Sue Oh, Steven D Charlier, Michael K Mount, and Christopher M Berry. The two faces of high self-monitors: Chameleonic moderating effects of self-monitoring on the relationships between personality traits and counterproductive work behaviors. *Journal of Organizational Behavior*, 35(1):92–111, 2014.
20. Bennett J Tepper, Michelle K Duffy, Christine A Henle, and Lisa Schurer Lambert. Procedural injustice, victim precipitation, and abusive supervision. *Personnel Psychology*, 59(1):101–123, 2006.
21. M. R. Barrick, G. L. Stewart, M. J. Neubert, and M. K. Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, 83(3):377, 1998.
22. Blaine Landis. Personality and social networks in organizations: A review and future directions. *Journal of Organizational Behavior*, 37(S1), 2016.
23. James Coleman. Relational analysis: the study of social organizations with survey methods. *Human organization*, 17(4):28–36, 1958.
24. Henk Akkermans. Renga: a systems approach to facilitating inter-organizational network development. *System Dynamics Review*, 17(3):179–193, 2001.
25. Eva Schiffer and Jennifer Hauck. Net-map: collecting social network data and facilitating network learning through participatory influence network mapping. *Field Methods*, 22(3):231–249, 2010.
26. A Zilli, F Grippa, P Gloor, and R Laubacher. One in four is enough–strategies for selecting ego mailboxes for a group network view. In *Proc. European Conference on Complex Systems ECCS*, volume 6, pages 25–29, 2006.
27. Peter A Gloor. What email reveals about your organization. *MIT Sloan Management Review*, 57(2):8, 2016.
28. Francesca Grippa, Antonio Zilli, Robert Laubacher, and Peter Gloor. E-mail may not reflect the social network. In *Proceedings of the North American Association for Computational Social and Organizational Science Conference*, pages 1–6, 2006.
29. Richard Borovoy, Michelle McDonald, Fred Martin, and Mitchel Resnick. Things that blink: Computationally augmented name tags. *IBM Systems Journal*, 35(3.4):488–495, 1996.
30. Jonny Farringdon, Andrew J Moore, Nancy Tilbury, James Church, and Pieter D Biemond. Wearable sensor badge and sensor jacket for context awareness. In *Wearable Computers, 1999. Digest of Papers. The Third International Symposium on*, pages 107–113. IEEE, 1999.
31. Alex Sandy Pentland. Automatic mapping and modeling of human networks. *Physica A: Statistical Mechanics and its Applications*, 378(1):59–67, 2007.
32. Daniel Olguín Olguín and Alex Sandy Pentland. Social sensors for automatic data collection. *AMCIS 2008 Proceedings*, page 171, 2008.
33. Lynn Wu, Benjamin Waber, Sinan Aral, Erik Brynjolfsson, and Alex Pentland. Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an it configuration task. 2008.
34. The big five project personality test. https://www.outofservice.com/bigfive/, 2000. Accessed: 2017.
35. Moral foundations questionnaire. https://www.yourmorals.org. Accessed: 2017.
36. Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
37. Siena. http://www.stats.ox.ac.uk/~snijders/siena/siena_r.htm. Accessed: 2017.
38. Tom AB Snijders, Gerhard G Van de Bunt, and Christian EG Steglich. Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60, 2010.
39. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. $D^3$ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.
40. D3plus. https://github.com/alexandersimoes/d3plus. Accessed: 2017.

# APPENDIX I

### PRE-STUDY SURVEY

**Q1.1** Which 3 people in the group do you think exchanged most **emails** with you during **last week/month/year**?

**Q1.2** For the 3 names you provided above, what do you think are their 2 most active hours in a day in sending emails?

**Q1.3** Which 3 people in the group do you think had most **face-to-face** communication with you during **last week/month/year**?

**Q1.4** Which 3 people in the group do you think had most communication with you combining **all channels** (emails, f-f, Slack, WhatsApp, etc) during **last week/month/year**?

**Q2.1** Who do you think are the 3 most central people in the group? Please rank them in descending order. Here central means having strong email communication with members in the group.

**Q2.2** Which 3 people in the group do you think are best connected with the most central person (the 1st name in last question)?

**Q3.1** How many emails do you think you sent to **all your contacts** during 2017?

**Q3.2** How many emails do you think you sent to **all the other members** during 2017?

**Q4.1** How many college majors are people in the group from? Can you list them on descending order of the number of people from the college major?

**Q4.2** How many countries are people in the group from? Can you list them on descending order of the number of people from the countries?

**Q5** Who in the group do you think is most similar to you in personality?

**Q6** To which 3 people in the group do you respond the fastest?

**Q7** Which 3 people in the group do you think respond to others fastest overall?

**Q8** Rank people in the user study based on their strength of email communication (more emails) with you. On the top should be the person who exchange more emails with you.

### MAIN SURVEY

**Q1.1** Who are the 3 people in the group exchanged most **emails** with you during **last week/month/year**?

**Q1.1.1** Are the three names the same with your answer of Q1 in the pre-study survey? If no, what do you think made you feel the names in Q1 in the pre-study survey exchanged most emails with you?

**Q1.2** For the 3 names, what are their 2 most active hours in a day in sending emails?

**Q2.1** Who are the 3 most central people in the group? Please rank them in descending order. Please rank them in order. Here central means having strong email communication with members in the group.

**Q2.2** Which 3 people in the group are best connected with the most central person (the 1st name in last question)?

**Q3.1** How many emails do you think you sent to **all your contacts** during 2017?

**Q3.2** How many emails do you think you sent to **all the other members** during 2017?

**Q4.1** How many college majors are people in the group from? Can you list them on descending order of the number of people from the college major?

**Q4.2** How many countries are people in the group from? Can you list them on descending order of the number of people from the countries?

**Q5** Who in the group do you think is most similar to you in personality?

**Q6** To which 3 people in the group do you respond the fastest?

**Q7** Which 3 people in the group do you think respond to others fastest overall?

**Q8** What did you learn about the team network? Anything interesting you found in your team network?

### POST-STUDY QUESTIONNAIRES

#### 5-Point Likert-Scale Questions

1. I feel MITeams is easy to learn.
2. I feel MITeams is easy to use.
3. I feel MITeams reflects my interactions with people in the group accurately.
4. MITeams helps me learn more about the communication and dynamics in the group.
5. I learn more about people in the group after the user experiment.
6. I'm willing to use MITeams in the future to analyze team communication network.
7. I would like to recommend MITeams to a friend.

#### Open-Ended Questions

1. Do you take account of hours when you send emails?
2. What is the main resistance for you to use the tool? (anonymity, loading time, difficulty to use, other privacy issues, etc)
3. For the data you see in MITeams, are there anything you would like to anonymize or remove from your team?
4. Do you feel the demo before the study useful? Why?
5. Do you think you are able to explore the tool without watching the demo?
6. What do you think MITeams would be useful for in a

company?
7. What do you think MITeams would be useful for in your group?
8. Which features (can be but not limited to any details in visualizations, functions, subpages, etc) in MITeams do you feel most useful?
9. Which features (can be but not limited to any details in visualizations, functions, subpages, etc) in MITeams do you feel interesting to explore?
10. Which features (can be but not limited to any details in visualizations, functions, subpages, etc) in MITeams do you feel difficult to use/explore?
11. What are other features you would like to see in MITeams?