2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Email Conversation Network Analysis: Work Groups and Teams in Organizations

Sarka Zehnalova, Zdenek Horak
Inflex, s.r.o.
Olomouc, Czech Republic
Email: {zehnalova, horak}@inflex.cz

Milos Kudelka
VSB - Technical University of Ostrava
Ostrava, Czech Republic
Email: milos.kudelka@vsb.cz

***Abstract*—Email communication is a source of important information, much of which is at first sight hidden. This paper presents an analytical tool that was created to analyze the deeper relationships in the email data. Those include relationships based on an interaction of multiple users in a team. The analytical methods proposed and described in this paper are based on two factors. The first factor is the context, which is a group of multiple users in combination with terms used in the communication. The second factor is the time interval in which the communication was conducted. Based on these factors, we analyze the conversations that take place and get results that are in several different forms presented to the users. The paper presents methods for weighting conversations, users and relationships, as well as methods for finding communities associated with the specified context. Additionally, the concept of the explorative user interface is introduced.**

***Keywords*—social network analysis, email networks, message threads, team communication**

I. Introduction

Teams in many organizations nowadays constitute major units for dealing with various tasks. Teams consist of members with high task interdependency and shared common goals [1]. Consequently, much research has focused on models of team performance, teamwork or team effectiveness resulting in a science of teams. Some of the factors that influence team performance are communication structure, work assignments, workload, task type, interdependency, etc. [2].

Our research focuses specifically on communication structure, workload and roles that we extract from emailing activities of several persons in an organization. The focus is on project/team-oriented long-term communication. To gather information necessary for analysis we use mailboxes which are a great source of data since they include information about users (sender, recipient, CC, and BCC), mails (subjects, message bodies, attachments) and sequences (answers, forwarded emails). Every team, formed formally or informally, has throughout the time a different workload. Long-term email communication in the team is reflected in the mailboxes of individual team members. Analysis of email data can reveal interesting and, at first sight, unseen information. Much of semantics is hidden in the data. This semantics can be visualized in different ways to obtain a result that can then be used for decision support.

The first step is to obtain the data. It is necessary to choose the team members, from whose email addresses the data for analysis will be extracted. After that, a simple procedure starting with extraction and ending with exploratory analysis of the communication, is applied.

During the analysis of team's performance, the demandingness of individual conversations, user's share in them and important terms (keywords) are measured and assessed. The result is a detailed knowledge of the community, especially the major users and strong ties between them, and also about important terms in communication. This knowledge is used to produce an understandable visualization of the community and the details of its communication.

This paper presents a novel approach to email mining with a focus on teams and exploration using the TeamNet application. We evaluate the method in a case study of real-life email communication of Inflex LLC company.

The rest of this paper is organized as follows. In Section II, we discuss related work. The proposed method is presented in Section III. The system overview is described in section IV. In Section V, we focus on real-world case example. Section VI concludes the paper.

II. Related Work

The major tasks that involve investigation of email involve spam detection, email categorization, contact analysis, email network property analysis and email visualization [3]. Our work covers the areas of email network property analysis and visualization. We will focus on previous work in those areas.

A. Email Network Analysis

Techniques and models of social network analysis (SNA) have been applied to discover the relationships between people, groups, and organizations from email networks. Many studies [4], [5], [6], [7], [8] used Enron email corpus as a dataset due to the lack of large public email corpus.

Diesner et al. [9] constructed directed graph from relations (sender to recipient) where edges were weighted by the cumulative frequency of emails exchanged in a time range, then applied SNA methods (network measures, centralities, etc.). Rowe et al. [8] constructed an undirected graph with edges representing communication between two accounts and introduced a social score measure for a user (based on number of emails, response score, clique scores, centrality measures etc.) which is then used to automatically reconstruct the social hierarchy. Chapanond et al. [5] used network metrics and

ASONAM '15, August 25-28, 2015, Paris, France
© 2015 ACM. ISBN 978-1-4503-3854-7/15/08 $15.00
DOI: http://dx.doi.org/10.1145/2808797.2808862

spectral analysis to study both directed and undirected email graphs constructed by changing the value of the threshold (e.g. minimum number of emails). The difference between such studies and our approach is in the construction of the communication network (relations through conversations), also in our approach this network is only one of the views derived from email activities. Furthermore, our tool is exploratory, it allows the user to refine the context, i.e. view of the data, by adding relevant persons or topics to the context, or changing the time span.

Some models were designed to work with the language content or topics besides relations. McCallum et al. [6] presented the Author-Recipient-Topic (ART) model that discovers discussion topics (using probabilistic language modeling) in a corpus of messages, extended also to the Role-Author-Recipient-Topic (RART) model to capture multiple roles (by clustering sets of correlated topics). In our approach, only email subjects are processed to extract keywords that are then associated with a conversation, but roles are defined independently of those keywords.

Stuit and Wortmann [10] use the business process modeling language to visualize the process. They construct message threads (conversations) in the same manner used in our approach (utilizing message header fields) and also use the conversation as a basic unit of observed communication. However, we do not analyze each conversation as a unit but use a set of them together to assess the social aspect of communication.

B. Email Visualization

Visualization tools can be used to analyze an individual email account or email archives. NodeXL [11] tool includes an email import tool where analysts can generate email networks based on the sender/recipient fields of an email corpus. It generates direct reply networks [12] and SNA metrics (in-degree, out-degree, centrality, etc.) of those networks can be used to identify important people or social roles (by creating ego-networks of each contributor). NetLens-Email [13] is a system designed to support exploration (using queries) of the content-actor network in large email collection providing a visual representation of the data over several attributes (e.g., distribution of the number of emails by a time period), or distribution of people by status (e.g. sender vs. receiver). Themail [14], [15] is another tool for visualization of relationships (interaction histories) in email archives. It is designed for an individual's email archive, extracting keywords from the content of exchanged messages and visualizing how they change over different time periods. Our approach provides ego-centered visualization of email network around selected user, but the main focus is on teams, by selecting a group of users, the visualization provides the view of email network for this group.

III. Background

In this section, we define the key concepts that we use in our email analytics tool.

A. Conversations

Definition 1. (*Conversation*): Conversation is a set of emails which:

- begins with a single email
- has at least two emails where senders are different
- other emails are either replies or forwarding of some email from that conversation

Conversations can contain multiple branches of different length. A conversation can be more or less taxing for the participants. From our observations, we estimated several factors that influence the difficulty of conversations. Conversation is demanding when:

- has many emails in total
- has many emails in a longest thread (branch)
- goes through many days
- has many emails in days when users are conversing
- has many emails where senders are different

Definition 2. (*Conversation demandingness*): Let C be the conversation, M be the total number of emails in a conversation, x is the logarithm base empirically set to 1.4, $Mmax$ be the number of emails in the longest branch, N be the number of days in which participants conversed, $Nmax$ be the maximum number of emails in one conversation day, S be the number of senders in the conversation. Then $D(C) = D(M, Mmax, N, Nmax, S), D(C)$ is *demandingness* of conversation C, defined by Equation 1.

$$D(C)=\sqrt{\frac{(\log_x(M-1))^2+(N-1)^2+Nmax^2+F^2}{4}}\cdot\frac{Mmax}{M} \tag{1}$$

Remark 1. For further calculations, we use the normalized conversation demandingness $Q(C) \in [0, 1]$, $Q(C) = D(C)/R$, R is the maximum value of $D(C)$.

B. Roles

User's roles are detected from his/her participation in a conversation; it can be either active or passive. On that account we divide the roles into two groups, we call the initiator, solver and co-solver *active* roles and the rest *passive* roles.

- *Initiator*: the sender of the first mail in the conversation
- *Solver*: the recipient of the first mail (not a copy) who sends at least one mail during the conversation
- *Co-solver*: the sender of some mail (not a copy) in the conversation
- *Invited*: the recipient of the first mail (not a copy) in the conversation
- *Co-invited*: the recipient of some mail (not a copy) in the conversation
- *Notified*: the rest

Definition 3. (*User weight in conversation*): Let C be the conversation, U be the user in the conversation C, All be the number of emails in a conversation, $From$ be the number of emails in which the user U is the sender, To be the number of emails in which the user U is the receiver, $Copy$ be the number of emails in which the user U is in CC or BCC (and not in To or $From$). Parameters f, t, c give importance to each component; their sum equals 1. Then $w(U, C)$ is *user weight in the conversation*, defined by Equation 2.

$$\begin{aligned} w(U,C) = f \cdot F + t \cdot T + c \cdot C \\ F = \frac{From}{All}, f = 0.7 \\ T = \frac{To}{All}, t = 0.2 \\ C = \frac{Copy}{All}, c = 0.1 \end{aligned} \tag{2}$$

Because weight $w(U, C)$ does not reflect the intensity of the conversation from a global perspective (different user weight in easy and challenging conversation), we use normalized conversation demandingness $Q(C)$ to convert the user weight according to the demandingness of conversation, so $w_n(U, C) = Q(C) \cdot w(U, C)$.

Definition 4. (*Term weight*): Let C be the conversation, T be the term, then $w(T, C)$ is the weight of term T in conversation C, $w(T, C) = Q(C)$ if conversation C contains term T, $w(T, C) = 0$ if conversation C does not contain a term T.

Remark 2. Based on statistical analysis, we found out that the vast majority of conversations does not change terms in the email subject. Therefore, the weight of each term in the conversation equals normalized demandingness of the conversation.

Definition 5. (*Relationship weight*): Let C be the conversation, U_1 and U_2 be users in the conversation C, then $w(U_1, U_2, C)$ is the weight of relationship of these users in the conversation, defined by Equation 3.

$$w(U_1, U_2, C) = MIN(w(U_1, C), w(U_2, C)) \tag{3}$$

C. Context and community

A context is a group of specified users and terms (from email subjects). The specified context is a subject for further analysis where the joint communication of specified users concerning the occurrence of specified terms is examined. Key results of this analysis are:

1) **Community** consisting of users specified as the context and users who interacted together with them. Users in this community are divided into two subgroups. The first subgroup consists of active users who have been in the community communication senders of emails. In the second subgroup are passive users who have been in the community communication only recipients.
2) **Community terms** that occurred in the community communication. Terms are also divided into two subgroups. In the first subgroup are terms that occurred together in the communication of all community users. The second subgroup contains terms that have been used only in the communication of some of the community users.
3) **Community conversations** that constitute the communication between the community users.

Finding the community is based on context parameters (a group of users, a group of terms, or a combination of users and terms). A community based on the specified group of users is constructed as follows:

1) Enter the users and mark them as *context users*.
2) Find all conversations in which are all *context users* together in an *active* role. Mark these conversations as *community conversations*.
3) To the *context users* add users who are in an *active* role in at least one community conversation. Mark these users as a *community*.
4) Find all terms that appear at least in one community conversation. Mark these terms as *community terms*.

A community based on the specified group of terms is constructed as follows:

1) Enter the terms and mark them as *context terms*.
2) Find all conversations in which all *context terms* appear together. Mark these conversations as *community conversations*.
3) Find all users who are in *active* role in at least one community conversation. Mark these users as *community*.
4) To the *context terms* add terms that appear in at least one community conversation. Mark these terms as *community terms*.

Community based on the specified combination of users and terms is constructed as follows:

1) Enter the users and mark them as *context users*.
2) Enter the terms and mark them as *context terms*.
3) Find all conversations in which are all *context users* together in an *active* role and where all *context terms* appear together. Mark these conversations as *community conversations*.
4) Find all users who are in an *active* role in at least one community conversation. Mark these users as a *community*.
5) To the *context terms* add terms that appear in at least one community conversation. Mark these terms as *community terms*.

IV. System Overview

In this section, we describe basic steps of our approach starting with an individual's mail account and ending with calculated statistics, roles, etc. The process includes data collection, pre-processing of data, aggregation of data from multiple user accounts and data analysis, see Fig. 1.

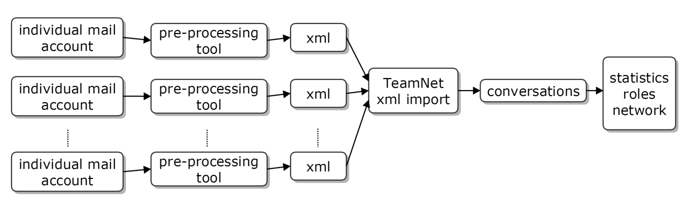


Fig. 1: System overview.

A. Pre-processing

To extract data from email account the system is using a separate application that contains the connectors to email clients (MS Outlook, Thunderbird) or connects via IMAP or Gmail. The output of the application is an XML file (dataset) ready for further use. The pre-processing tool is using the header fields [16] (Message-ID, In-Reply-To, References) to reconstruct conversations (threads) and message subjects to extract keywords (terms). Pre-processing doesn't utilize the message body nor attachments; the reasons are legal restrictions on access to personal information. After the pre-processing of emails from an individual email account produced XML file can be then imported into TeamNet application. In case of a single file from just one person, the resulting views provide insight only from this ego-centric perspective. The intended usage for TeamNet is with a set of multiple XML files collected (in the organization) from different persons.

B. Data import

Collected XML files are processed together and aggregated while removing duplicities, into one set and stored in SQL database. SQL database enables effective and quick work with five types of entities: users and the relationships between them, with the keywords (terms) from communication and relationships between them, and with conversations. Specially designed algorithms work with these entities.

C. Analysis method

Conversations are weighted according to their demandingness in terms of time cost or complexity. User roles are derived from the type of user's participation in the conversation (see Section III-B), we also calculate the user role's ratio (weight) in the conversation. The calculation of weights is based on definitions and equations in Section III.

D. User interaction

After the pre-processing, the work with TeamNet application starts with a query that defines the context. It can be a single user (term), a group of users or a combination of user(s) and term(s). Part of this query is also the selection of desired time span. Based on the selected context, the conversations are filtered, and user (term) weights and user role ratios for this context are aggregated. System algorithms automatically detect the characteristic of the team (defined by context) and provide different ways of visualization. The flow of user interaction is depicted in Fig. 2.

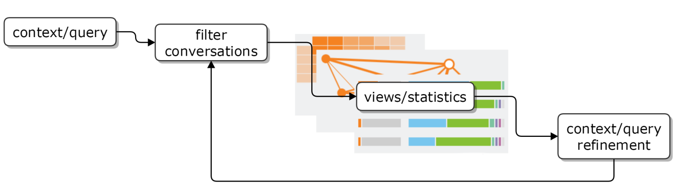


Fig. 2: TeamNet usage after importing email accounts.

The views include the charts (basic statistics), activity heat map, graph of main domains, user network, term cloud/network, communication roles, etc. Every view offers a different visualization of the same context. The views themselves are interactive and allow for further refinement of the context. By changing the time span, it is possible to observe how roles in the team change over time, or how the surrounding network evolves. The information from the views about the team's workload may help to manage resources more efficiently.

Remark 3. Context (users, terms and the time span) may be chosen randomly. The purpose of the analysis of the context is a selection of a maximum number of key users and terms that are most frequent in the communication within the specified context. In our case, this value is set to 50. Therefore, the output of the analysis is at most 50 weighted users and terms and, weighted relationships between them. As described in Section III-C, the basis for identifying users and terms is common conversations. Over the time, however, common conversations are changing, depending on the occurrence of users and terms in conversations. At the end of the calculation, algorithms only select a limited number of users and terms and then filter conversations. Selected users and terms must be understood in a context of the selected set of conversations and the specified time span.

V. Case study - Inflex LLC

Inflex LLC is a small computer software company founded in 2004, developing desktop, web, mobile and Smart TV applications tailored to customers' needs. The company provides help desk support to their customers through phone and particularly email. Number of employees is around ten and number of projects is around fifty. All projects depend heavily on email communication. Due to the small number of employees the hierarchy structure is flat, most of them are developers, one businessman/manager plus one more executive.

A. Inflex LLC dataset

Email data were collected from six individual accounts (most included more than one email address per account), this team consisted of four developers, one businessmen and one executive. Altogether, $156,876$ messages from $16,740$ users (email addresses) in the period from $1.1.2010$ to $21.1.2015$ were analyzed. For details see Tables I and II.

TABLE I: Inflex LLC - Dataset Overview

	Count
Emails	**156,876**
Conversations	**18,833**
Conversation branches	**29,967**
Users (email addresses)	**16,740**

TABLE II: Inflex LLC - Dataset Details

	Count
Users who are in at least one conversation	**2,936**
Emails sent from someone from Inflex LLC	**64,076**
Conversations in which at least one user is from Inflex LLC	**16,880**
Conversations initiated by someone from Inflex LLC	**9,385**

B. Project example

The outputs from TeamNet application are always context-specific. Due to the absence of one comprehensive report with statistics about the whole collection (that could be evaluated), we evaluate outputs using one selected company project, IntraDoc, as an example. IntraDoc is a company's long-term project with several slightly different instances deployed for several customers. So the keyword 'intradoc' is selected as the context term and the time span is the whole interval (2010-2015).

First view is charts, see Fig. 3, where the most important persons are listed on top. The system correctly identified the manager (Hovorka), head developer (Spacil) and executive (Kudelka) as being most involved. Key topics associated with this project revolve around words like 'version, problem, new application, modification', etc. Key persons for all roles are also listed.

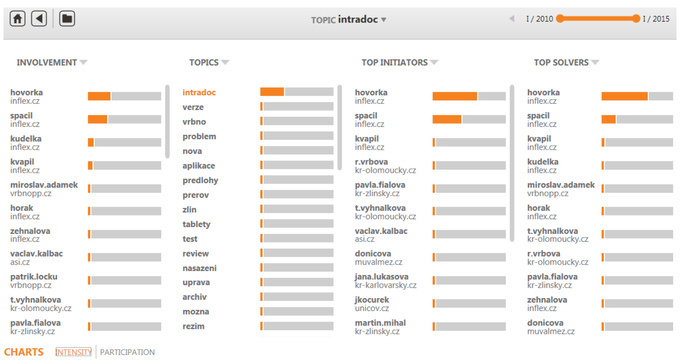


Fig. 3: TeamNet chart - major team members for IntraDoc

Next view is a network of all users associated with the IntraDoc project and their relations, see Fig. 4. Around two prominent nodes, manager (Hovorka) and developer (Spacil), are nodes (customers) with direct link only to those two. It is clear that the rest of the company employees rarely communicate directly with customers, which corresponds to the reality. This network as an output can be further analyzed using SNA methods, i.e. simple community detection algorithm [17] was implemented to detect interconnected subgroups, see Fig. 5, where all company employees are detected as one of those subgroups.

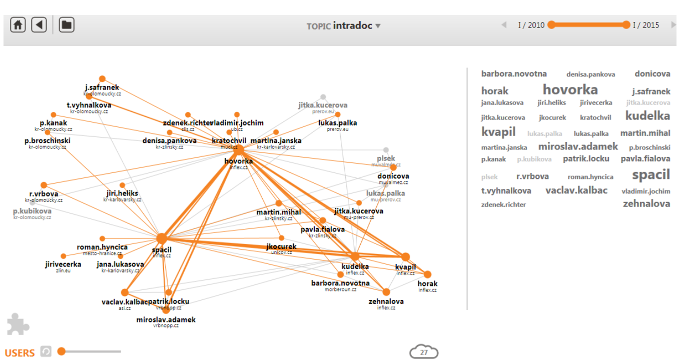


Fig. 4: TeamNet network - users relations in IntraDoc

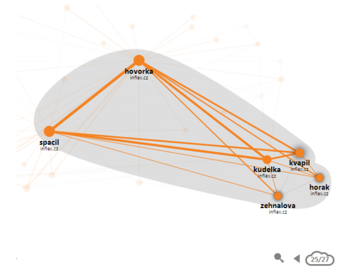


Fig. 5: TeamNet network - detected subgroup - Inflex employees

Another view in Fig. 6 displays the distribution of roles among the team members. We can see the manager and head developer on top again, now with additional information about their roles, and ratio of each role w.r.t. others. We see another developer (Kvapil) on 4th place (in involvement) who acts as 'solver', which reflects his real role on this project. He handled the data/business layer of the product.

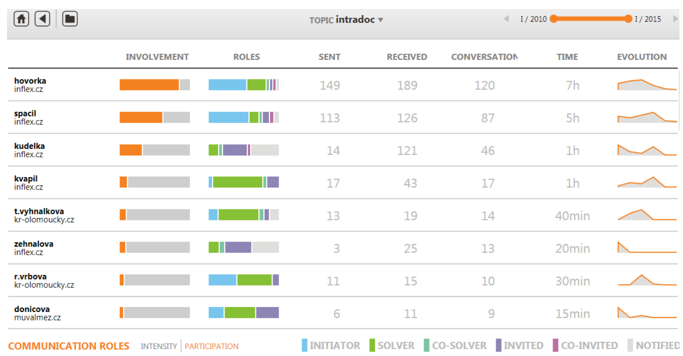
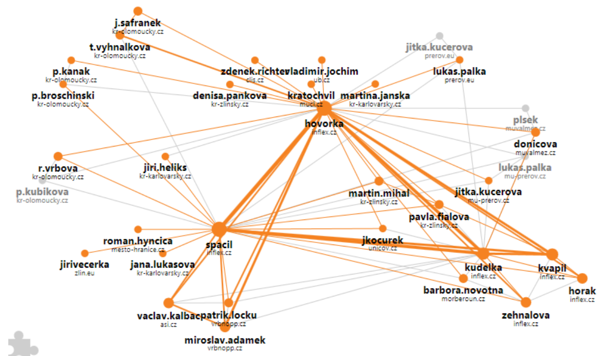


Fig. 6: TeamNet roles - team members distribution of roles in IntraDoc

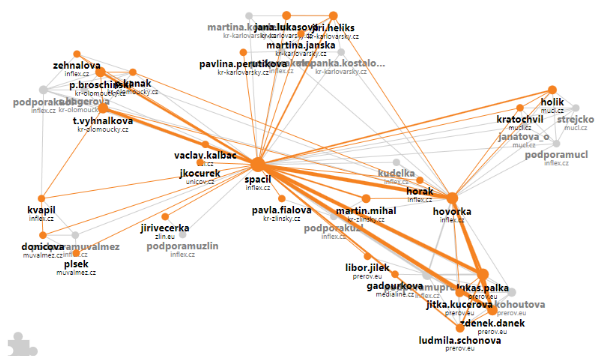
C. Project example with limited time interval

In next example, we use the same project but focus on the evolution of relationships and workload by changing the time span. We restrict it to the interval of only one year (I/2014 - I/2015). By that time, the system was deployed and running for years, so the team's workload is gradually declining. The communication structure between individual team members changed which can be seen in the communication network, see Fig. 7, where the central position of manager (Hovorka) in Fig. 7a is taken by the head developer (Spacil) in Fig. 7b, who by being at the support desk developed strong relations with customers. The communication structure also changed between organizations, see Fig. 8, where the partners are different in each interval.

The distribution of roles evolved in the same manner as the communication structure, see Fig. 9, where the most involved person became the head developer (Spacil) by being at the support desk, and the two passive addresses below him are support for customers, see Fig. 9b. Except for the manager (Hovorka) the rest of Inflex employees is not involved significantly anymore.

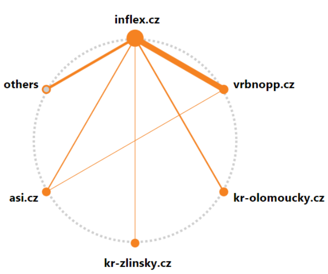


(a) I/2010-I/2015

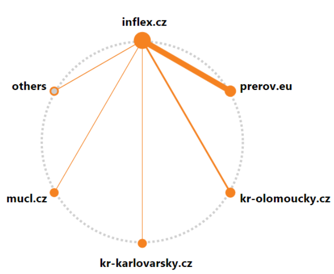


(b) I/2014-I/2015

Fig. 7: TeamNet network evolution



(a) I/2010-I/2015



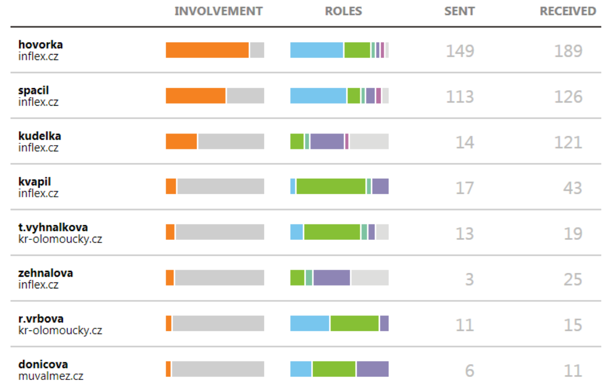
(b) I/2014-I/2015

Fig. 8: TeamNet organizations evolution

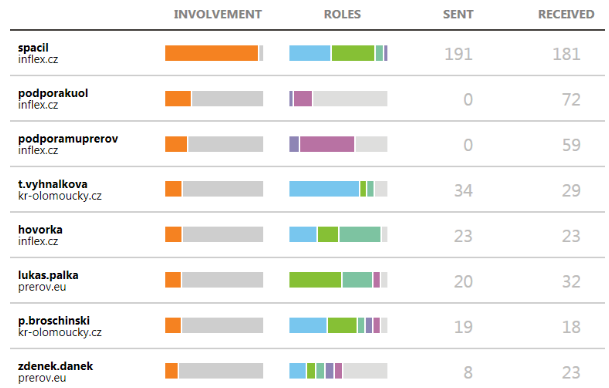
Remark 4. Take notice of the selection of time interval in the two examples above. We do not compare two different intervals of the same length; the second interval is part of the first (five years vs. one year). In the first case, we look at the project globally (using whole interval) and the importance of team members is assessed correctly (overall, the manager played the most important role). However, it would be wrong to draw conclusions from those charts about how the situation looks like at the present moment, as evidenced by the different results in the second example.

D. Accuracy

As described in Section III, the presented approach is based on measuring the demandingness of conversations. The calculated values are then used for weighting users, terms



(a) I/2010-I/2015



(b) I/2014-I/2015

Fig. 9: TeamNet roles evolution

and relationships between them. To assess the accuracy of calculations, we performed a simple experiment. We automatically detected the list of the forty most important users of the entire dataset over the time. Furthermore, we selected all the important projects that the company has worked on. Each of the six employees (their accounts were used for the analysis) completed a questionnaire and identified most important persons in the individual projects. Afterward, a list of all persons who were identified in at least one project (75 persons in total) was created. These persons, together with the automatically detected persons, were aggregated to the small experimental dataset (85 persons in total). We classified all persons from the questionnaires as positive and the others (10 persons) as negative. Then we took into account the first ten (to forty) automatically detected persons by weight and calculated precision and recall values. The results are shown in Table III.

TABLE III: Precision and Recall (the first n-selected persons)

Selected	TP	FP	TN	FN	Precision	Recall
10	10	0	10	65	**1.00**	**0.13**
20	15	5	5	60	**0.75**	**0.20**
30	21	9	1	54	**0.70**	**0.28**
40	30	10	0	45	**0.75**	**0.40**

E. Evaluation

By a combination of two factors, the context, and the time interval, we get a set of conversations. With deeper analysis of this set, we can detect community (participants of the conversations) and topic (terms of conversations) that characterize the behavior of the team at that time. There are thousands of such sets of conversations defined by the specified context. The communities matching these contexts vary in size from two participants to hundreds of participants. One person or a single term may appear in hundreds of conversations, and dozens or even hundreds found communities of different sizes. This effect implies the importance of explorative approach, allowing refine the selected context interactively. For reasons of efficiency of calculations, the data are preprocessed in the system, so the exploratory interface responds very quickly.

The system was piloted in two other organizations with teams of three to eight members (the first organization specialized in project management, and the second was the university). Despite skeptical expectations, the result for team members was surprising. Mainly for two reasons:

1) Outputs from directly specified context matched expectations.
2) The exploratory interface facilitated discovering situations that in detail showed surprising outcomes (e.g. low or excessive workload and interconnectedness of some people).

VI. Conclusions

In this paper, we proposed methods that are focused on the analysis of email communication of small teams. These methods were designed so that the basic unit of communication was the conversation. A weighing method was designed for those conversations, being also the basis for weighting users, terms and relationships in the communication. We also presented algorithms for finding communities specified by the context and selected time interval. All the procedures presented in the paper were used in the implementation of the analytical tool that was deployed in several organizations (working in project teams to a large extent) as a pilot project. This analytical tool is built on an exploratory user interface that provides a combination of different views of the analyzed data. Despite the very good feedback, pilot users missed automatic interpretation for displayed outputs. These are largely sociological tasks that are not simple. The second problem is the ability continuously to process incoming messages and to provide more detailed information about the development of communication. In future research, we will focus on both directions.

VII. Acknowledgments

This work was supported by the EUREKA project no. E!7182 (Optimising of working team by using a sw for analysis of the social and professional relations in the company networks) and by SGS, VSB-Technical University of Ostrava, under the grant no. SP2015/146.

References

[1] J. L. Dyer, “Team research and team training: A state-of-the-art review,” *Human factors review*, vol. 1983, pp. 285–3, 1984.

[2] E. Salas, N. J. Cooke, and M. A. Rosen, “On teams, teamwork, and team performance: Discoveries and developments,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 540–547, 2008.

[3] G. Tang, J. Pei, and W.-S. Luk, “Email mining: tasks, common techniques, and tools,” *Knowledge and Information Systems*, vol. 41, no. 1, pp. 1–31, 2014.

[4] J. Diesner and K. M. Carley, “Exploration of communication networks from the enron email corpus,” in *SIAM International Conference on Data Mining: Workshop on Link Analysis, Counterterrorism and Security, Newport Beach, CA*. Citeseer, 2005.

[5] A. Chapanond, M. S. Krishnamoorthy, and B. Yener, “Graph theoretic and spectral analysis of enron email data,” *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pp. 265–281, 2005.

[6] A. McCallum, X. Wang, and A. Corrada-Emmanuel, “Topic and role discovery in social networks with experiments on enron and academic email,” *Journal of Artificial Intelligence Research*, pp. 249–272, 2007.

[7] P. S. Keila and D. B. Skillicorn, “Structure in the enron email dataset,” *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pp. 183–199, 2005.

[8] R. Rowe, G. Creamer, S. Hershkop, and S. J. Stolfo, “Automated social hierarchy detection through email network analysis,” in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007, pp. 109–117.

[9] J. Diesner, T. L. Frantz, and K. M. Carley, “Communication networks from the enron email corpus it’s always about the people. enron is no different,” *Computational & Mathematical Organization Theory*, vol. 11, no. 3, pp. 201–228, 2005.

[10] M. Stuit and H. Wortmann, “Discovery and analysis of e-mail-driven business processes,” *Information Systems*, vol. 37, no. 2, pp. 142–168, 2012.

[11] D. Hansen, B. Shneiderman, and M. A. Smith, *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann, 2010.

[12] D. L. Hansen, B. Shneiderman, and M. Smith, “Visualizing threaded conversation networks: mining message boards and email lists for actionable insights,” in *Active Media Technology*. Springer, 2010, pp. 47–62.

[13] H. Kang, C. Plaisant, T. Elsayed, and D. W. Oard, “Making sense of archived e-mail: Exploring the enron collection with netlens,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 4, pp. 723–744, 2010.

[14] F. B. Viégas, S. Golder, and J. Donath, “Visualizing email content: portraying relationships from conversational histories,” in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 979–988.

[15] M. E. Joorabchi, J.-D. Yim, M. E. Joorabchi, and C. D. Shaw, “Enron case study: Analysis of email behavior using emailtime.” in *IEEE VAST*, 2010, pp. 235–236.

[16] P. Loshin, *Essential email standards: RFCs and protocols made practical*. John Wiley & Sons, Inc., 1999.

[17] S. Zehnalova, M. Kudelka, and V. Snasel, “Comparing two local methods for community detection in social networks,” in *Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on*. IEEE, 2012, pp. 155–160.