E-mail May Not Reflect The Social Network

Francesca Grippa [1], Antonio Zilli [1]
[1] *e*-Business Management Section, ISUFI, University of Lecce, Italy
francesca.grippa@ebms.unile.it, antonio.zilli @ebms.unile.it

Robert Laubacher [2], Peter A. Gloor [2]
[2] MIT Center for Coordination Science
rjl@mit.edu, pgloor@mit.edu

Abstract. This paper aims to demonstrate that ties obtained by mining e-mails archives do not necessarily provide a complete and realistic approximation of interactions by other communication media. The results of our project indicate that factors such as *co-location* and the *nature of working relationships* influence the preference for face-to-face interactions and chat. Peers who are co-located and who are working on the same project are more likely to use synchronous media rather than e-mail. By adopting a social network approach we compared the complete networks implied by four different media (e-mail, face-to-face, chat and phone), measuring network properties like betweeness centrality, density, and core/periphery structure, in order to identify at what extent the network implied by e-mail differs from the network based on other communication media.

Key words
interpersonal communication, network topology, intra-organizational networks, internet-based surveys

1. Introduction

Previous studies demonstrated that e-mail interaction can be a good indicator of knowledge flows within organizations [i] [e]. The main strength of this technology is the plenty of data easily accessible to researchers regarding personal communication. E-mail parsing tools allow researchers to gather data for Social Network Analysis (SNA) by using email archives [d] [k]. Although this data gathering method is faster and easier than traditional survey tools, it might provide a view of the network that is different than traditional tools. Adopting a methodology that minimizes the possible biases of survey technique, we will try to find out how to improve the effectiveness of e-mail parsing through a comparison of networks implied by other communication channels. This study is not interested in deepening technical problems related to e-mail parsing method: different data formats to be managed within the SNA tool, privacy issue, difficulty to obtain e-mailboxes from different organizations. Our research interest is more focused on finding biases of e-mail in representing the overall communication ties within a community of co-workers.
The main goal of this study is to understand whether email parsing provides the same view of the network that traditional SNA surveys do. We identified the main differences between network implied by e-mail and networks based on other media, so to understand the potential biases of email and see if they can be corrected so to take advantage of the speed and ease of email parsing and still achieve an accurate view of networks. Firstly, we measured the overall communication patterns of a community of students and researchers with reference to traditional and electronic media. Then we used some networks' properties like density, betweenness centrality [Freeman, 1979] and core/periphery structure [Borgatti, Everett, 1999] as further indicator of possible misalignments between e-mail and the complete knowledge flow.

2. Research Site Description

Our research site has been the e-Business Management School, a graduate school of e-business management that both educates students and undertakes research projects. Its organizational structure is characterized by an emphasis on flexibility, informal contacts and communication supported by a variety of traditional and electronic media. Deadlines are associated with software demonstration, courses presentation, and project submissions to external

committees. The research activities are project-oriented and members worked by objectives (project deliverables, software development). We monitored the communication patterns of 25 members (50% of the entire population) who worked in two on-campus buildings. In the *student building* (64% of sample) students and programmers worked in four different laboratories, divided in project teams and assigned to junior faculty in charge of implementing project activities. In the *techie building* (20%) a staff of technicians worked assisting students and researchers in implementing hardware and software solutions. Students shared two adjacent offices and worked full time in *team* attending the same research laboratories and classes, differently from technicians and junior faculty, who were the most mobile members and worked more individually. Economic and engineering expertise was equally represented in the sample (48% and 52%).

3. Hypotheses

In a setting characterized by co-located team members, mapping only the exchange of e-mails might be not a sufficient indicator of social ties. The following hypotheses are defined:

Hypothesis 1: *The communication pattern of e-mail differs significantly from patterns of other media.*

Hypothesis 2: *Network implied by e-mail and network implied by face-to-face, chat and phone together are different in terms of density, betweeness centrality and core/periphery structure.*

We examine these hypotheses by exploring the main similarities and differences among networks, by taking as point of reference the *complete network*, defined as the network implied by four media taken together: e-mail, chat, phone, face-to-face interactions (both scheduled and unscheduled).

4. Data Collection Methods

This study has been conducted through a combination of different sources of evidence: web surveys and mining e-mailboxes archives, in order to obtain structured data on the number of times peers communicated with each other using different media. To parse emails we adopted TeCFlow, a software suite that analyzes unstructured information and automatically generates interactive movies of communication flows [d]. Survey results were tabulated using an application written by an eBMS programmer. Data was parsed and imported into MySQL databases.

In the web survey respondents have been provided with a roster of all the people working in the organization. They were asked to report on a web page (linked to a database) "the number of times they communicated by phone, by in-person interactions, and by chat, the same day they received the questionnaire", only with reference to work-related issues. In order to increase the reliability of data collected, we avoid to use the Name Generators technique, as literature on free recall have found that respondents forget between 1% and 90% of the people they are asked to list [h]. Survey was administered once per week for 7 weeks on a daily basis, and e-mail archives were referred to the entire survey period (September 8th -November 4th 2005). Asking people to recall only interactions occurred the same day of the survey is an approach that sensibly reduces the risk of huge errors in reporting communication ties [a]. Indeed, forgotten network ties would make recalled network data incomplete and possibly distort measurement of various structural characteristics of social networks. Questionnaires have been sent in different week days, so to avoid an over-representation of days with a regular high interaction (i.e. Friday meetings).

The weakness of survey regarding "poorly constructed questions" [Yin, 1994] has been managed by reporting complete instructions and explanations within the personalized e-mails of invitation, as well as on the survey's home page and during informal discussions. Furthermore, because of the presence of international students within the community, the first two questionnaires have been provided in a twofold linguistic version so to avoid any linguistic misunderstanding.

Finally we involved three researchers, selected for their role as Project Managers and Education Managers and for their tenure (more than 4 years in the School), in focused structured interviews and face-to-face discussions in order to add a subjective dimension to the definition of the eBMS' social capital. The open one-to-one discussion took as starting point their responses to a brief questionnaire sent by email, where they had to assess each actor listed in roster according to the relative level of *trustworthiness* within the team, *prestige* among peers and *contribution* to the team performance, using a Likert scale (1=very low, to 5=very high). In this study we considered as global index of *perceived contribution* the mean value obtained by each actor in all the three above mentioned sub-indices.

5. Data Analysis

Survey data was normalized according to members' response rates (7,6,5 times): the number of time people communicated by each mode was multiplied by 1, 1.4 or 1.16, so to obtain comparable data on the basis of 5 working days. The approximation of daily interaction was based on two different normalization methods for e-mail and survey data. The average daily interaction has been calculated on 45 working days (9 weeks period) for e-mail and on 7 days for the survey data. Structured data about actors' interaction collected from survey and email archives was managed through valued adjacency matrices, where each cell contained values of communication ties assumed as non-directional. Since we were interested in any connection between members, the directions of ties are ignored, so either a out-tie or an in-tie constitutes a tie between two actors. The TeCFlow tool was used to export adjacency matrices and to automatically calculate network properties. The strength of the relation was measured by the number of interactions between actors by each medium, so that the more interaction between actors, the closer the nodes representing vertices will be placed in the graph [d].

6. Findings

6.1 Overall Communication Patterns

Preliminary evidences indicate a strong predominance of face-to-face and chat communication ties, as well as a scarce use of e-mail. Of the 190 communications/day between the 25 members of community, about 96% were represented by in person interactions, chat and phone. E-mail represented only 4% of the daily communication pattern, although the ability to parse email by computer leads to an "amplification of the e-mail signal", as we observed 22% of communication ties by e-mail over the complete interactions (table 1). By considering the average daily interactions, we found that 77% of the total daily ties was explained by face-to-face ties, 16% by chat messages, 3% by phone calls. These results indicate how co-workers greatly use synchronous media, strongly prefer voice based media like face-to-face that have a high presence, and then use a synchronous, text-based, and more "trackable" medium like chat. This preference for supports results of previous studies [i] [g] on media use: in person interaction, more than e-mail, allows to convey meta-information and to avoid any equivocality, leading to a more effective promotion of cooperation.

	Communication ties observed (# and %)					Daily communication ties (# and %)				
	E-mail		Other media		Complete	E-mail		Other media		Complete
Total communication ties	352	22 %	1272	78 %	1624	8	4 %	182	96 %	190
- Inside groups	210	15 %	1170	85 %	1380	5	3 %	167	97 %	172
- Between groups	142	58 %	102	42 %	244	3	17 %	15	83 %	18

Table 1. Communication ties inside and between groups [1]

We considered the "location of co-workers" as a factor influencing their communication patterns. In our study 64% of members was located in the *student building*, 20% in the *Techie building* and 16% worked *off-site* part of the time. We found that peers working in open-space laboratories located at different floors tend to be internally connected more through a digital technology like chat than by e-mail (2.7 chat/day and only 0.3 e-mail/day). We adopted the index O/E (ties by other media divided by e-mail exchanged) as a preliminary indication of the comparative rate of occurrence of e-mail in the complete network: the highest this index, the more incident will be other media over email. The communication pattern of the complete network is characterized by a prevalence of other media, used 22 times more than e-mail.

The O/E index indicates that e-mail was used on average 1,5 times more than other media for communication between groups and less for interactions within the same co-located group. Most of the e-mail exchanged every day was sent and received by people working abroad part of the time or by technical staff who usually moved from one building to the other. Then, of the overall 2016 face-to-face communication observed, 60% were reported by people with economic background, and 40% by members with an engineering background, who were also connected more by e-mail (59%) than by in-person interaction (41%). This result confirms what literature on communication between functional groups [Moenaert, Souder 1990] has pointed out: R&D teams appreciate written communication because of its higher credibility, whereas marketing prefers face-to-face communication due to its higher comprehensibility.

1 In table 1, "communication observed" are values normalized according to the response rate and refer to 9 weeks period. "Daily communication" are normalized according to 45 days (for e-mail) and 7 days (for survey).

The involvement on a specific project is a further factor which can help to identify the reason behind the great use of in person interactions. Part of our sample (35%) worked on two projects that were in their initial phase, and characterized by a scarcely structured knowledge base. Members of these projects strongly used face-to-face communication more than e-mail, while chat and phone were used similarly to e-mail. Then, 65% of sample was allocated on two mature projects characterized by a well codified and shared knowledge base. Even in this case we calculated the ratio "other media over e-mail" (O/E), which was equal to 8.5 and to 12 for initial project teams and to 1.6 and 2.3 for mature project teams, suggesting that in the planning phase of a project, members interact more frequently using informal channels, in order to generate a shared knowledge for future activities [Souder & Moenaert, 1992]. The knowledge they exchanged in this critical stage was mostly *tacit* and not codified in any support, so the *socialization* process [Nonaka, Takeuchi 1994] was facilitated by synchronous media, supporting the conversion of fluid knowledge and reducing any form of ambiguity and equivocality [i]. On the contrary, the peculiar activities of a mature project team suggest to adopt media enabling a formal diffusion of the results and an exploitation of knowledge generated within the group. Thus, formal media like e-mail might be the most appropriate to allow the externalization of discrete and codified knowledge.
The predominance of synchronous media is also explained by the type of task co-located workers were asked to accomplish: prepare classroom presentations, design courseware, or developing software. All these tasks required a high degree of interaction and a quick response from colleagues. The need to overcome linguistic misunderstanding, given the variety of languages spoken in the laboratories, may represent an additional reason for our results, as synchronous media represent an important vehicle of trust.

6.2 Networks Comparison through Density, Betweenness Centrality and Core/Periphery structure

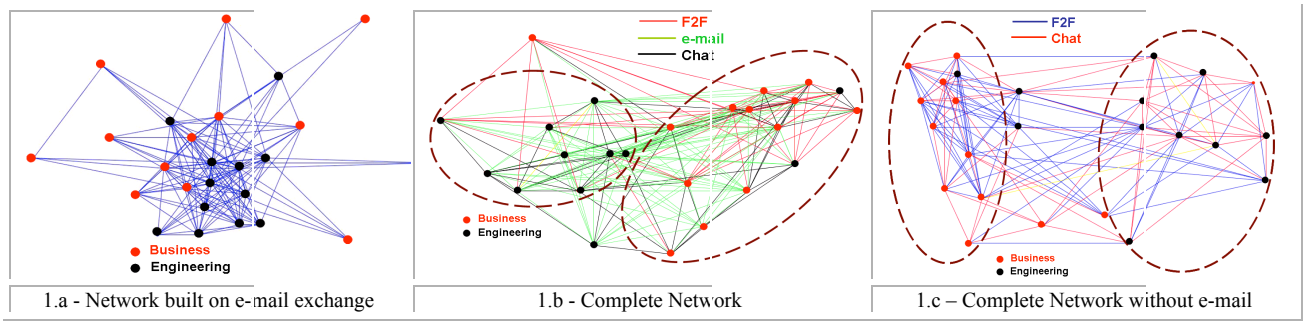
The first way to compare networks has been to look at their density, intended as the number of peers who communicated by e-mail and by other media over the maximum number of 300 pairs "respondents-correspondents". Our results indicate how every medium provided an essential contribution to get an overall definition of how formal and informal knowledge flows within the group. In order to understand how each medium helped to define properties of the whole network, we considered the density for each medium and measured how connections among pairs increased from one network to the other as percentage of the whole network. We found that e-mail participated with 72% of connections to define the density of the complete network, followed by in person contacts (62%), chat (38%), and phone (24%). Then we combined one by one various networks and considered the sequential improvement in the approximation of the whole network. The total edges of the network implied by merging *e-mail and chat* explained 85% of the overall network density. It means that by combining the two electronic media we get a better approximation of the complete information flows. E-mail network alone had the highest value of group density, mostly explained by the presence of a core group of techie actors dominating the communication, and it was the most indicative medium of the overall communication ties (72%), while phone is the least indicative (23%). Finally, the complete network's density was entirely described by the density of networks based *e-mail and face-to-face* together.

Type of network	# of nodes	# of edges	Group Betweenness centrality	Group Density	Core/Periphery Fitness Values	% Density [over the complete one]
e-mail	25	149	0,117	0,248	0,2823	71,9 %
f2f	25	130	0,113	0,216	/	62,6 %
chat	25	78	0,175	0,13	/	37,7 %
Phone	19	28	0,384	0,082	0,2452	23,8 %
chat + f2f + phone	25	138	0,0865	0,23	0,2046	66,7 %
chat + e-mail	25	175	0,046	0,292	0,2410	84,6 %
f2f + e-mail	25	207	0,0275	0,345	0,2125	100 %
Phone + e-mail	25	157	0,0862	0,262	0,2711	75,9 %
f2f + chat	25	137	0,0947	0,228	/	66,2 %
Complete	**25**	**207**	**0,0275**	**0,345**	**0,2125**	

Table 2. Properties of networks and their combination.

Table 2 shows how complete network was characterized by a lower level of group betweenness centrality with respect to e-mail network. This decentralized communication pattern allowed members to finish complex tasks in less time than groups with centralized communication networks [c]. Thus, a more egalitarian communication pattern

seems to be possible when knowledge flows across multiple channels. A similar result emerges from core/periphery structure, whose Fitness Values indicate how well the observed data approximated an ideal core/periphery structure [d] and allow to identify the presence of dense core and sparse periphery. The highest fitness value for e-mail based network, together with the absence of such a structure for chat and face-to-face ones, might be associated to a situation of teams engaged in non-routine, complex tasks, where core-periphery structures might impede teams' effectiveness, by marginalizing information coming from peripheral members. So, networks implied by face-to-face or chat indicate that people with diverse expertise were freely interacting for solving a given problem [c].
The graph of e-mail network taken alone [graph 1.a] had a centralized, dense structure, with a core occupied by programmers, technicians and junior faculty, who supported peripheral students located at the periphery (less than 1 year of seniority and an economic background). Differently, complete network [graph 1.b] shows a decentralized structure, with two sub-groups composed by members with engineering expertise (left side), and people with economics skills (right side). Four of the most central people, two for the business and two for the engineering side - acted as gatekeepers between different skilled peers. The complete map allowed to visualize edges between co-located peripheral actors involved on the same projects, which were missing in the e-mail graph.



1.a - Network built on e-mail exchange | 1.b - Complete Network | 1.c – Complete Network without e-mail

Graph 1. Graphs of different networks implied by different media

Complete network without e-mail [graph 1.c] reveals a separation between co-skilled people as in the complete network. It also points out the presence of a structural hole between these groups, bridged by a master student with an engineering background who is the most central actor in both face-to-face and chat networks, and who worked in the school far before her master colleagues. She acted as gatekeeper connecting students and junior faculty located in different buildings, and she received a high score in terms of assessed contribution (*perceived contribution*=4.2). Her "bridging role" could have not been recognized by simply looking at her betweeness centrality value in e-mail network (9th position).

6.3 Local Betweenness centrality

By comparing actors' betweenness centrality in various networks, we aimed at find out which actors were prominent and particularly visible to the other actors in the complete network and which in the e-mail one. We assumed that people who were central in almost all the networks had more positional advantage over others who were central only in few networks. The focus on betweenness more than degree centrality is due to our interest in identifying actors who play an "intermediary" role across all the media, who may have a low degree centrality, but also the most access over information flows [b]. By considering the top 15 most central actors and the 10 most peripheral in the complete network, we found that network based on face-to-face had the highest level of similarity with complete network both in terms of central people (11 of the top 15 were the same) and peripheral ones (6 of the last 10 were the same most peripheral actors). Thus, face-to-face approximated complete network more than other media.
Moreover, people with engineering skills were mainly central in the e-mail network (60% of the top 15 were engineers), while business and technical expertise was equally represented in the complete network's top 15 central (50%). Network implied by e-mail overrates not only actors with technical expertise, but also the centrality of people who got a low score of *perceived contribution*, and underestimate the centrality of people playing an "information broker" role in the complete network. By looking at few examples, a first misalignment in terms of actors' betweeness centrality within e-mail and complete network is provided by a PhD student who is the most central within the complete network, but ranked 18th in e-mail network. A second evidence is given by a student scarcely involved in the school activity, ranked 10th in e-mail network, 21st in the complete one, and 25th according to the

perceived contribution index. The opposite case in which an actor might be considered very central according to e-mail network, although peripheral in the whole network, is represented by a Project Manager, ranked 6th in the complete network but only 22nd in e-mail one. These evidences of misalignment in actors' centrality suggest that by simply mapping information exchange by e-mail might lead managers to assume a wrong centrality for actors, whose role in the network depends also on informal relationships with peers.

7. Conclusions

This study have identified some biases of e-mail mining method, such as: a) the overestimation of communication between peers with technological expertise and their central role within the network, b) a lack in representing ties between peripheral co-located team members, c) an underestimation of individuals playing a gatekeeper role, d) the risk to interpret the communication pattern as dominated by a core group, although the democratic exchanges described by the complete network may lead to the opposite conclusion. An integrated network based either on *e-mail and chat* or *e-mail and face-to-face* within the same software package may help decision makers to make different and broader managerial implications from social network maps, by capturing both the codified and tacit knowledge. They might discover how people create a trust environment through face-to-face or chat; recognize the importance of information brokers and *informal leaders*, increase the effectiveness of the follow up phase of SNA in figuring out the real dynamics of group communication.

The "daily elicitation" approach to web-survey we have followed in this study should be able to increase the correlation between self-report and real behaviour and might represent a good way to integrate e-mail mining methods with traditional survey technique. The first point in our future research agenda will be to investigate the biases of e-mail with reference to geographically dispersed groups.

Acknowledgement

We would like to thank Yan Zhao, the technical staff of the eBMS, and the survey participants who also gave their consensus to using their e-mailboxes. Without their collaboration, this research would have not be possible.

References

[a] Allen T.J., "Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization" Cambridge MA, MIT Press, 1984.

[b] Cross R., Borgatti S. P., Parker A. "Making Invisible Work Visible: using social network analysis to support strategic collaboration", California Management Review vol. 44, no. 2 winter 2002.

[c] Cummings Jonathon N., Cross Rob, "Structural properties of work groups and their consequences for performance" Social Networks 25, 197–210, 2003.

[d], Gloor P.A., Laubacher R., Dynes S.B.C., Zhao Y. "Visualization of Communication Patterns in Collaborative Innovation Networks: Analysis of some W3C working groups", ACM CKIM International Conference on Information and Knowledge Management, New Orleans, Nov 3-8, 2003.

[e] Kossinets G., Watts D.J., "Empirical analysis of an evolving social network", Science, Vol. 311, January 6, 2006.

[f] Haythornthwaite C., Wellman B., "Work, friendship, and media use for information exchange in a networked organization", Wiley & Son Journal of the American society for information science, 1101 – 1114, 1998.

[g] Jensen C., Farnham S.D., Drucker S.M., Kollock P., "The effect of communication modality on cooperation in online environments", 1999.

[h] Marin, A. "Are respondents more likely to list alters with certain characteristics? Implications for name generator data" Social Networks, 26, 4. 289-307. 2004.

[i] Trevino L.K., Webster J., Stein E.W., "Making connections complementary influences on communication media choices attitudes and use", Organization Science, Vol. 11, n.2, pp. 163-182, Mar. Apr. 2000.

[j] Tyler, J. Wilkinson, D. Huberman, B. A. "E-mail as Spectroscopy: Automated Discovery of Community Structure within Organizations", HP Laboratories, 2003.

[k] Wellman, B., "Computer networks as social networks: collaborative work, telework, and virtual community". Annual review of Sociology, vol. 22, 213-238, 1996.