Three essays on Organizations Networks and Knowledge

Anna Velyka

PhD Supervisors: Prof. Marco Guerzoni, Luca Bolognini

Vilfredo Pareto Doctorate in Economics

XXXI Cycle

Contents

1 Introduction

The resource-based theory of the firm that developed in the post-industrial economy defines knowledge as one of the most critical assets in the organization [Penrose, 1959, Teece, 1981, Grant, 1996, Davenport and Prusak, 1998]. Employees quit and retire, taking with them decades of company knowledge. Even if they train their replacements before leaving, they are never able to pass on everything they know because some knowledge is tacit and hard to communicate. Unintentional loss of knowledge may also arise as a result of increased reliance on technology (storing knowledge on local drives (hard disk or e-mails). Moreover, knowledge stored locally is not fully available to others. An effective knowledge management system reduces the costs of human and technological inefficiencies by making company knowledge more accessible and accurate. The death of the distance brought about as the result of technological advancement in the field of communication is the "most important force shaping society" [Cairncross, 1997]. The speed and quantity of information transfers arising as a positive externality of these changes disrupt modern-day organizational arrangements. The challenge ahead of corporate management is not only to embrace the transformations brought by technological disruption but also to transform it into improvements in workplace solutions.

Traditionally interventions in human resource practices rely on data about individual characteristics and quantitative indicators of performance [Holton III and Yamkovenko, 2008]. Employee's organizational role is defined through the set of individual skills and characteristics. As working arrangements shift away from traditional hierarchies to more collaborative practices based on the team collaborations a new level of capital emerges - relational or social. Social capital reflect the value of the employee from a different standpoint.

Social interactions play an indispensable role in developing participative, inclusive and learning-oriented organizational culture [Heydebrand, 1989, Fulk and DeSanctis, 1995]. The growing importance of social capital established communication processes as the mean through which organizations are designed and sustained. However, corporate management often overlooks social practices be they connected with internal or external communications. Conventional metrics used in the human resource management prioritize quantitative and financial outcomes [Boyd and Gessner, 2013]. One of the possible reasons for the lack of recognition is data availability. The record keeping summarizing interaction activities in the organization can be time- and effort-consuming. What is more, the evaluation of communication practices requires advanced analysis techniques to extract possible metrics connected to organizational performance.

Organizational literature summarizes three possible techniques applied to gather the data about collaboration arrangements and identify key performers: observations, interviews, and questionnaires [de Toni and Nonino, 2010]. Such data collection expects active participation of the research subject, which allow for maximum precision in the definition of analysis objective. These data mining techniques offer a flexibility of focus allowing to summarise insights about personal impressions about other actors, established working practices, expectations, etc. However, a self-reported methodology is labor intensive and may lead to a series of limitations: response rate sensitivity, bias, record inaccuracy [Yamkovenko and Hatala, 2015]. New perspectives in organizational science are opening up with the research shifting to digital data sources. The direct interaction with the research subject is no longer crucial as data is mined from the variety of ICT tools utilized to support information-sharing activity among employees (e.g. e-mail, telephones, calendars, collaboration platforms, intranets, and wikis) [Yuan et al., 2013]. This is a passive way of information extraction. Its application provides the researcher with the advantage of higher data scale and longitudinal perspective. The costs of data gathering and preparation are significantly lower than for the active methods due to the electronic form in which data is stored [Tashiro et al., 2010]. Two significant obstacles on the way of passive approach integration in the organizational network analysis is a privacy concern and high computational requirements.

The intangible nature of social asset imposes restraints on measuring and analyses techniques. Traditional statistical methods may appear inefficient in developing a methodological framework for making inferences about the structure of communication between employees. The theoretical basis of network science combined with data on human resources offers a sophisticated solution for granting a numerical expression to socialization practices [Monge and Contractor, 2003, Cross et al., 2002].

Since 1930's Social Network Analysis has emerged as an approach for studying communication structures [Freeman, 1978]. SNA allows stepping beyond the borders of the tree-like hierarchical graph, providing an alternative view of individual connections within the organization. Not all the links connecting employees are visible and implicit. SNA is a flexible method for representing countless interactions hidden within organizational structures [de Toni and Nonino, 2010] through its broad range of algorithms and metrics. One of the prominent organizational scientists Rob Cross refers to such structured quantitative and qualitative analysis of informal arrangements as Organizational Network Analysis. "ONA can provide an x-ray into the inner workings of an organization - a powerful means of making invisible patterns of information flow and collaborations in strategically important groups visible" [Dulworth and Dulworth, 2008].

This thesis will advance three arguments to support the inclusion of communication practices analysis in organizational research. I will argue that studying organizations through the prism of the social network analysis enables analytical assessment of interaction practices and broadens understanding of employees performance. I present two real-life cases for the application of Organizational Network Analysis within the realm of collaboration practices. This work has led to two important conclusions: practical and methodological. The practical conclusion states that visualizing human resources through the network perspective provides a quantitative expression of collaboration practices. This approach should be embarrassed by companies to balance two potentially conflicting aims effectively: fairly reward individual participation for decreasing undesired turnover, and at the same time develop recognition mechanisms for acknowledging the role of teamwork. Conducted research reveals that central people in the organizations are not necessarily the ones on the top of the hierarchy. Identifying informal leaders and ensuring their efficient physical allocation should improve knowledge flows, idea sharing, and decrease coordination costs. Moreover, identification of informal leaders who accumulate high share of social capital gives an early indication of possible overload and consequent burnout.

Understanding how employees' positions in the communication networks reveal hidden patterns in knowledge sharing practices across-departments allowing to increase returns from the application of available knowledge and expertise within the organization. The analysis of organizational knowledge flows reflected through employees' communication patterns is important to prevent the loss of human and information resources.

The focus of research and data availability dictates the choice of methodological approach for the analysis of organizational interactions. A methodological conclusion reached in this work states that basing the analysis solely on the centrality statistics calculated in the static network significantly reduces the predictive power of the model. Communications are continuous by nature. As such, social network must be periodically revisited or measured at least twice to ensure a good fit and higher predictive power of the model. Using longitudinal data extracted from digital communication tools can significantly enrich the interpretation of centrality statistics and provide an alternative view of individual performance.

Both active and passive Organizational Network Analysis should be applied to ensure adequate human resource management. Surveying organizational actors allow to accurately identify employees that are critical to the organization and thus serve as a measurement for the behavior that organizations are looking to replicate. Passive ONA should be integrated to enable a longitudinal perspective on organizational interactions. This research provides indications for switching to the continuous perception of performance based on the people-centered metrics. Implementing Organizational Network Analysis allows identifying a path that an employee takes to reach effectiveness and maximum involvement in the organizational realm.

References

N. Boyd and B. Gessner. Human resource performance metrics: Methods and processes that demonstrate you care. *Cross Cultural Management*, 20(2):251–273, 2013.

F. Cairncross. *The death of distance: How the communications revolution will change our lives.* Harvard Business School Press, Boston Massachussets, 1997.

R. Cross, S. P. Borgatti, and A. Parker. Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, 44(2):25–47, 2002.

T. H. Davenport and L. Prusak. *Working knowledge: How organizations manage what they know.* Harvard Business Press., 1998.

A. F. de Toni and F. Nonino. The key roles in the informal organization: a network analysis perspective. *The Learning Organization*, 17(1):86–103, 2010.

M. Dulworth and M. Dulworth. *The connect effect: Building strong personal, professional, and virtual networks.* Berrett-Koehler Publishers, 2008.

L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

J. Fulk and G. DeSanctis. Electronic Communication and Changing Organizational Forms. *Organization Science*, 6(4):337–349, 1995.

R. M. Grant. Toward a knowledgeâĂŘbased theory of the firm. *Strategic management journal*, 17 (S2):109–122, 1996.

W. V. Heydebrand. New organizational forms. *Work and occupations*, 16(3):323–357, 1989.

E. F. Holton III and B. Yamkovenko. Strategic Intellectual Capital Development: A Defining Paradigm for HRD? *Human Resource Development Review*, 7(3):270–291, 2008.

P. R. Monge and N. S. Contractor. *Theories of communication networks.* Oxford University Press, USA., 2003.

E. T. Penrose. *The Theory of the Growth of the Firm.* New York: John Wiley., 1959.

H. Tashiro, J. Mori, N. Fujii, and K. Matsushima. Email Network Analysis for Organizational Management. (June 2009):958–963, 2010.

D. J. Teece. The market for know-how and the efficient international transfer of technology. *The Annals of the American Academy of Political and Social Science*, 458(1):81–96, 1981.

B. Yamkovenko and J. P. Hatala. Methods for Analysis of Social Networks Data in HRD Research. *Advances in Developing Human Resources*, 17(1):40–56, 2015.

Y. C. Yuan, X. Zhao, Q. Liao, and C. Chi. The use of different information and communication technologies to support knowledge sharing in organizations: From e-mail to micro-blogging. *Journal of the American Society for Information Science and Technology*, 64(8):1659–1670, 2013.

The more you ask, the less you get: the negative impact of collaborative overload on performance

Anna Velyka [1,3] and Marco Guerzoni [1,2]

[1] Despina, Big Data Lab, Department of Economics and Statistics "Cognetti de Martiis"

[2] ICRIOS, Bocconi Univerisity

[3] Aizoon

Abstract

This paper is about the possible negative impact of collaboration within an organization on the performance of employees. With the rise of participatory culture and developments in communications technology, management practices require greater conceptual awareness about possible outcomes of increased organizational interconnectivity. While there exists a sound theoretical basis for possible burdens brought by collaborative overload, the literature has few empirical observations and verified models. We address this gap by developing a methodological framework for the identification of organizational actors at risk of operational capacity overload. Drawing on social network analysis as the widely applied approach for the estimation of employees' involvement in the information exchange networks, this paper describes potential personal and organizational causes leading to the emergence of overload. Relying on primary data gathered through a survey conducted among employees in an Italian insurance company, a testable model for overload detection is constructed.

This research suggests that active collaborative activity does not cause a decrease throughout every aspect of performance. We confirm previous findings indicating that expertise sharing depends on a few key players who take core knowledge assets upon themselves and thus run higher risks of exposure to overload. We also find that factors such as the nature of shared information and transmission channels have a significant influence on employees' performance. The results and the modeling technique are put forward so that they can be implemented in an organizational setting to improve information exchange practices and prevent the loss of high performing employees.

Keywords: organizational communication; employee performance; collaborative overload;

1 Introduction

The increase in intraorganizational communications facilitates knowledge diffusion and makes it easily accessible and retrievable. For all the advantages workplace collaboration brings, it can be a double-edged sword: active promotion of teamwork and peer support make reliable individuals more visible for colleagues and place an extra burden on communication activities. Cross and Gray [2013] and Cross et al. [2016] defined it as collaborative overload.

Inasmuch as communication plays a fundamental role in the day-to-day coordination of organizational activities, it might create severe bottle necks within the organization (ibid.). Collaborative interactions among employees is an underlying mechanism for defining the right person to execute particular steps and deliver an efficient organizational response to the challenges imposed by the turbulent dynamics of knowledge-intensive processes [Di Ciccio et al., 2012]. Based on hierarchical attributes and job titles, traditional reward structure deprived of the means to account for the collaboration as input and thus fails to reveal a full corporate value of an employee. Broader application of social network analysis provides the means to overcome the methodological challenge presented by the intangible nature of interactions and to quantify its potential contribution.

While communication has continued to gather momentum in the evaluation of organizational performance, the scientific community has warned of the possible negative externalities of intensive or ill-organized collaborative practices: more intensive intervention to boost organizational interactions are often costly to develop and may lead to unequal resource distribution, collective rivalry, and unethical behavior [Adler and Kwon, 2002]. Much-vaunted supportive culture may result in highly connected organizational community thereby hindering the individual performance of the actors overloaded by communication interaction and causing individual performance to decline [Cross and Gray, 2013, Oldroyd and Morris, 2012]. The concept of overload appears multifaceted in the academic literature. Different authors have invoked it as an umbrella term accounting for a decrease in information processing activity [Eppler and Mengis, 2004], intra-role conflict [Jones et al., 2007], and excessive collaboration demands [Cross and Gray, 2013]. Despite different phrasing, researchers agree on the negative influence this constraint inflicts on individual and organization performance. The extensive theoretical discourse originating from information processing, psychology, and organizational studies still lacks a corpus of empirical evidence due to the difficulty in identifying data sources and methods [Eppler and Mengis, 2004].

The present paper calls into question the notion of collaborative overload and empirically tests its effect on employee's performance. We propose to define collaborative overload as the point at which the performance of an employee declines as the result of the excessive time he or she spends elaborating incoming requests from colleagues for expertise sharing. We define performance as the ability of an actor to satisfy the demand for information.

This paper relies on the data obtained by a survey conducted in a big Italian insurance company to analyze the modes of interactions among employees. The survey methodology offers a practical way to assess the influence of organizational and personal characteristics on the efficiency of employees interactions. We construct an ordinal logistic regression to account for employees and tasks diversity and estimate the conditional probabilities of a higher level of efficiency in information sharing. Our research supports the hypothesis that social embeddedness of actors in knowledge exchange practices does not follow a normal distribution and key expertise holders are potentially exposed to negative externalities.

Thus, the present paper introduced quantitative methods to capture various organizational and personal factors affecting negatively individual processing capacity [Eppler and Mengis, 2004] and revealed particular aspects of performance that lead to overload [Oldroyd and Morris, 2012]. We contribute to the existing literature on the topic by providing an empirical analysis of collaboration pattern in the knowledge-intensive industry to capture possible negative externalities of ill-organized collaborative activities. Understanding the effect of collaboration on employees' performance is of great importance for effective human and knowledge resource management and in particular for devising operative decision-making processes.

2 Theoretical framework

The link between communication effectiveness and organizational performance has been the subject of inquiry for decades [Downs, C. W., Clampitt, P. G., Pfeiffer, 1988, Redding and Tompkins, 1988]. However, the literature initially overlooked the reciprocity of the impact between organization and individual, while managers understand intuitively and try to assess it, despite the difficulty in defining communication's contribution to the overall performance owed to an elusive nature of social contact.

Later on, teamwork and peer support became a precondition for establishing effective organizational culture [Tjosvold and Tsao, 1989] and also various research praised management intervention in intraorganizational collaborative practices, anticipating positive externalities for knowledge transfer, employees personal development and innovation diffusion [Goldhaber and Barnett, 1988, Downs, C. W., Clampitt, P. G., Pfeiffer, 1988, Tjosvold and Tsao, 1989, Dooley, 2002]. Relationship graphs of employees became a widespread technique to describe organizational interactions [Freeman, 1978, Ibarra, 1993, de Toni and Nonino, 2010, Oldroyd and Morris, 2012, Abbasi et al., 2014, Yamkovenko and Hatala, 2015]. By unveiling seemingly elusive personal networks, it is possible to highlight the role of collaborations for knowledge exchange [Granovetter, 1983], to describe the importance of brokerage roles for access and control of novel information [Burt, 1992], and, eventually, to analyze the impact of graph structure on performance [Tsai and Ghoshal, 1998]. More recent works have taken a step back shifting the focus to the explanation of network formation and the effect that network structure may have on organizational performance [Monge and Contractor, 2003].

Actors inside the organization are bound by a precise hierarchical structure [Adler and Kwon, 2002], which ensures the exchange and transfer of knowledge resources [Ansari et al., 2012] and facilitates access to colleagues who have the desired expertise [Andrews, 2010]. However, during the exchange of information a relational dimension also plays a critical role since it accounts for normative behavior based on trust, obligations, and expectations [Lee and Jones, 2008]. Analysis of the relational dimension reveals the willingness of an individual to prioritize the requests of colleagues while setting aside individual tasks and thus favoring collective goals [Lazarova and Taylor, 2009].

The evolution of such relational structures in the organizational setting may lead to the establishment of an efficient performance pattern or contribute to the development of a highly resistant system [Fukuyama, 2002]. However, social capital is prone to produce negative as well as positive externalities on different levels relating to individual, group, organization performance [Coleman, 1988]. While the positive externalities of social capital management are heavily studied, academic and business literature often overlooks the other side of the coin [Adler and Kwon, 2002].

One of the pioneering suggestions about the negative externalities of excessive communication originates in the early work by Wiio [1978], where the author invoked information processing capacity limit and questioned the universal nature of increased communication as a panacea. Individuals possess limited personal resources of time and energy. Given that organizational settings impose productivity driven attitudes and incentivize collaboration, individuals are encouraged to seek expertise from their colleagues to optimize the time required for decision-making and the quality of the decision itself [Borgatti and Cross, 2003].

The most commonly applied term to describe the conflict between individual processing capacities and processing requirements is information overload [Eppler and Mengis, 2004]. It originates from the inverted u-curve relationship between effective decision-making and information exposure. An upward trend is expected to last up until the point when the further delivery of new information fails to be integrated, and the performance of the actor starts to decline. Eppler and Mengis [2004] concluded that the overload results from the decrease in processing capacity of the receiver, his ability to integrate new information and to properly allocate time for processing. This research opts for the term *collaborative overload* due to the inability to fully account for the information flows employees may experience on a daily basis. Cross and Gray [2013] introduced the term "collaborative overload" to refer to possible unintended outcomes of increasing collaboration demand on high performing employees.

The combined power of cooperation between the organizational actors is higher than the total power of each separate individual. The synergy produced by the collective power [Jackson, 2010] also referred to as a teamwork [Cross et al., 2016] is proportional to the product of the time that each dedicates to cooperation. An actor addressing information request to the peer benefits from the time the two spend in cooperation. With each new request, the power of the original connection decreases as the demand for the personal resources of respondent increases. The more information requests an individual satisfies the less time he allocates to every cooperation [Jackson, 2010], leading to the emergence of negative externality from network centralization. As such, a mismatch in the personal motive and social capacity defines the cost of network inefficiency that becomes visible at the point when an increased number of connections results in more harm than benefit weakening the value of existing partnerships.

This paper seeks to address the setting strictly linked to the *professional service industry*, which is characterized by a limited number of key expertise holders [Oldroyd and Morris, 2012]. The object of analysis originates from relational data from a firm that reveals the structure of communication taking place between employees. The expectancy-value theory treats individuals as goal-oriented and defines two appraisals guiding their behavior: a belief that action will result in a conceived outcome and the evaluation of possible positive or negative degrees of this outcome [Palmgreen, 1984]. Thus, we assume that advice-seekers will be granting preferential treatment to those colleagues who are expected to deliver a more significant probability of a successful outcome. Some actors will be involved in knowledge sharing more than others and, therefore, restrict the decision-making process and concentrate essential knowledge resources in the minds of a few. The pressure of collaborative demand towards a limited group of individuals will be growing, requiring from them a higher level of information processing capacity, time management and balance of focus.

Previous works have primarily contributed to the theoretical analysis and conceptual discussion but failed to provide reliable empirical proof that understands circumstances under which the performance of focal actors decrease [Eppler and Mengis, 2004]. Several studies have proposed a testable model for the evaluation of possible effects of social capital abundance on employees performance [Speier et al., 1999, Allen, 2003, Eppler and Mengis, 2004, Oldroyd and Morris, 2012]. Collaboration in knowledge-intensive organizations is expected to have a cross-level pattern, neglecting hierarchical structure [Agneessens and Wittek, 2012]. Social network analysis reveals unobserved patterns of informal interactions among employees and shows informal focal actors [de Toni and Nonino, 2010] receive an abundant number of incoming information requests, which contribute to expertise accumulation and thus positively affect the chances of being approached again and thereby increasing their visibility in the organizational network [Tasi and Tsai, 2001]. Eventually, a growing frequency of incoming requests from colleagues conditions individual processing capacity of the top performers [Eppler and Mengis, 2004]. In their investigation into the social capital, Oldroyd and Morris [2012] described how a virtuous cycle of employees stardom might turn into a vicious cycle of overload. Cross et al. [2016] describes how "The avalanche of demands for input or advice [...] causes performance to suffer"(ibid. p.27).

This research picks up the debate by Adler and Kwon [2002] about the possible risks an excess in communication may create for individual and organizational performance by discovering how solidarity benefits embedded in the network of expertise-sharing restrict the time and cognitive capacity of givers thereby putting them under threat of a collaborative overload. We develop a model to test whether the employees' capacity to overcome overload differs [Oldroyd and Morris, 2012].

Specifically, we aim to verify whether there is a negative effect of excess of collaboration on the worker's performance in providing feedback to co-workers. Moreover, we investigate possible factors which might mediate this relationship, such as: personal factors (seniority, age, and gender), information characteristics (complexity of information), information technology (collaborative channel).

3 Data and Methodology

To test the hypothesis about the possible negative externalities of intensive collaboration, we collected data about organizational communications in a large Italy-based international insurance company. This professional service industry relies heavily on knowledge as an input and as the final output with the goal of customized solutions' delivery [Empson, 2001]. Organizations that produce tangible products (labor and capital-intensive industries) generate innovations by selecting the most lucrative options through internal R&D activities [Muller and Doloreux, 2009]. In contrast, knowledge-business intensive services prioritize organizational innovations that come as the result of a unique combination of technological and soft skills [Muller and Doloreux, 2009]. Thus, a professional service industry provides the setting where human capital is a dominant factor responsible for the generation of core knowledge assets [Oldroyd and Morris, 2012].

The evaluation of knowledge exchange through the prism of social network analysis has been widely used to map informal networks [Granovetter, 1983, Krachardt and Hanson, 1993, Abbasi et al., 2014], to identify influential actors involved in decision-making activities [de Toni and Nonino, 2010], to capture information exchange [Oldroyd and Morris, 2012], and to evaluate the organization's inclusion, connectedness and diversity [Yamkovenko and Hatala, 2015]. We define collaboration as the process in which organizational members initiate relational ties with colleagues in search of information resource required to maximize the efficiency of decision-making. Therefore, the basic unit of analysis in this study is intraorganizational communications for expertise sharing.

We made use of two sources of data. The primary data come from a survey providing information on informal interaction, task constraints, and a self-assessed quality evaluation on interaction. We merged this data with the second source summarising personal characteristics of the actors (age and gender) and seniority data. As for the survey, we used purposive sampling to identify the organizational actors of interest in our study. To identify key knowledge holders engaged in sharing professional expertise from the total population of employees we selected 303 professionals dealing with insurance claims settlement. Selected participants were asked to cite not more than five colleagues for each of the four major professional areas of the company. They were asked to concentrate primarily on the interactions and exchanges of information that happen beside the formal meetings, training courses and practices determined by the organizational rules [Krachardt and Hanson, 1993, Cross et al., 2002, Kaše et al., 2009, de Toni and Nonino, 2010, Oldroyd and Morris, 2012, Behrendt et al., 2014, Yamkovenko and Hatala, 2015]. By asking participants to focus on informal connections we recreated the exchange of the information not registered daily or established in the **ad hoc** documentation. The response rate was 79.6%, corresponding to 241 individuals generating 1437 citations.

Specifically, respondents were asked to mention 5 colleagues to whom they came for knowledge and information required to perform a certain task in the previous six months. For each person mentioned, respondents were asked to evaluate[1]:

1. The frequency of interaction on a scale from 1 to 4
2. The satisfaction of the answers received on a scale from 1 to 6 [2]
3. The type of information delivery requested in the question. The possible answer range from a very simple type of knowledge transfer such as "Simple instruction" up to a complex involvement of the co-worker such as "Step-by-step follow-up of the specific activity and verifying results" or "Practical demonstration with exercises and experiments on cases"
4. The communication channel, covering the usual mode of communications (email, telephone, personal meeting) [3]

Item 1 reflecting Frequency of collaboration between two individuals allows us to create a proxy for the burden of communication carried by different workers. One of the main factors in rising demand for an employee's processing capacity is the quantity and the frequency of the interactions required to satisfy the request for expertise sharing. We created the variable *Load*, which accounts for the intensity of collaborations experienced by the employee. Overload is expected to appear as the number of concurrent tasks and interruptions is rising and reducing the concentration of an employee, who experiences repeated context switching [Eppler and Mengis, 2004, Cross et al., 2016].

Item 2 attempts to measure the worker's performance. We opted for a subjective measurement of employees productivity through an individual opinion about request satisfaction. Thus,

[1] see Appendix 5 for a detailed description of the survey
[2] We chose an even scale of evaluation to avoid neutral responses as a way to skip the question.
[3] Other communication tools were not in use in organizational setting under study.

the variable of interest, $Efficiency$, measures the average efficiency of requests to collaborate as subjectively measured by the respondent of the survey.

We used item 4 to create the variable $Channel$ to control for the communication technology utilized to transmit information since it can create an additional constraint increasing the probability of overload occurrence [Eppler and Mengis, 2004].

Organizational environment witnessed a significant restructuring with the arrival of information technology revolution. In early work [Huber, 1990], and later [Dewett and Jones, 2001], suggested that information technology can be used as a significant factor influencing decision making quality and organizational performance. Indicators of proximity and frequency of interactions are also believed to facilitate the transfer of tacit knowledge [Polanyi, 1967]. The diversification of information transferring methods maximizes the scale and the frequency of communication. In this light, [Allen, 2003] anticipated potential risk of overload, due to the inefficient and unproductive use of information technology. By asking participants to indicate the means of communication used we determined the frequency and time required for interaction.

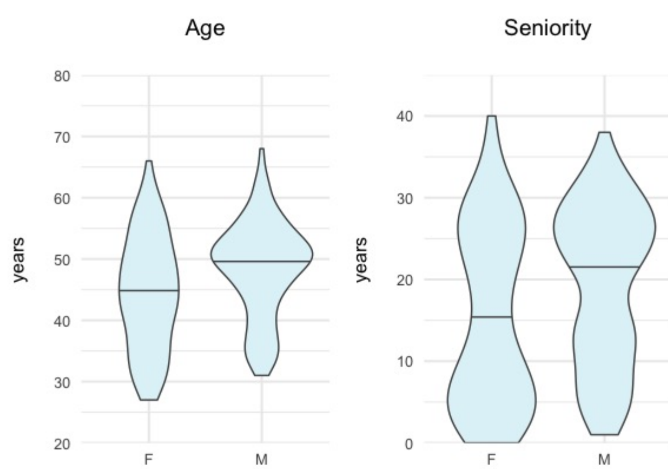
Item 3 controls for the complexity of the knowledge involved in the process. Schneider [1987] argued that the requirements for the information processing capacity are growing proportionally with the growth of the amount and the nature of information characterized by novelty, uncertainty, and complexity. Due to the higher requirements that these factors pose on individual cognitive capacity and time, they might also increase the possibility of overload occurrence. Due to statistical constraints explained in the next paragraph, we created the variable $Knowledge$ scaling down the initial six categories into two. The first category combines simple, one-time and more detailed instructions and explanations, to reflect information based on recurrent organizational practices. The second category, bringing together practical demonstration, analysis of the specific problem and step-by-step follow-up, accounts for more complex interactions with higher requirements to respondents personal resource.

Finally, organizational status is considered a powerful predictor of the strength of an employee's position [Krackhardt, 1990]. In their research on intraorganizational advice relations, Agneessens and Wittek [2012] assumed that the status and its strength could be defined by characteristics, leading to its emergence. The interest of this research is the social status, defined by competences and expertise, gained through the organizational learning process. Based on Berger's theory of expectation states, Agneessens and Wittek [2012] argued that regardless of being subjectively defined, a status can be derived from the general consensus. Employees' social worth and expectations are modeled through such variables as age and gender [Downs, C. W., Clampitt, P. G., Pfeiffer, 1988]. Thus, as controls, we included *Seniority*, *Age*, and *Gender* as individual characteristics that indicate informal leaders, social value and that might partially explain the centrality in the organizational network and the performance of the agents.

3.1 Descriptive statistics

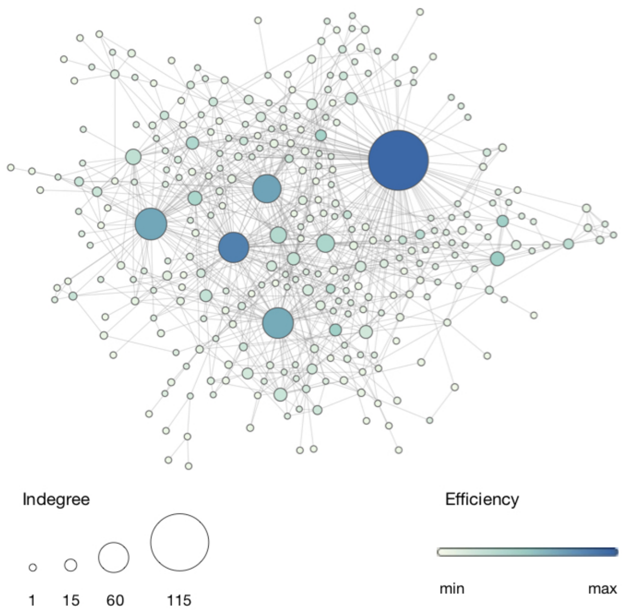
The sample consisted of 303 specialists engaged in the four principal areas of the company's expertise (anti-fraud, reserve management, civil liability insurance, property damages liquidation). 60% of the respondents were male. The age of the sample varied from 27 to 68 years, with a mean of 46 years. 35% of the workers had less than 10 years of working experience in the company, and 11% possessed more than 30 years. The majority of women have lower organizational tenure: 50% of females had less than 15 years of working experience inside the company, compared to 22 years for male (See Figure 1).

Figure 1: Distribution densities of Age and Seniority for female and male participants.



As result of the survey, we built a network of employees' interactions, accounting for 1437 directed edges. Figure 2 shows the network of informal communication based on the survey. The size of the nodes describes the number of contact requests and the color the average efficiency of interaction. As in most real networks, the degree distribution is highly right-skewed. The majority of agents in the organization network have low degrees while a small but significant fraction have an extraordinarily high degree. The figure clearly shows that the distribution of the collaborative burden is mainly on a few players. As expected, those key players exhibit a larger efficiency, since it is fairly reasonable to assume that they became central in the network precisely because of their reliability.

Figure 2: Network of collaboration and Efficiency



To account for collaboration activity, for each node we calculated the variable *Load* by summing the frequency of each incoming edges. This variable was unequally distributed over the sample of employees, suggesting the presence of loaded top performers in the organization.

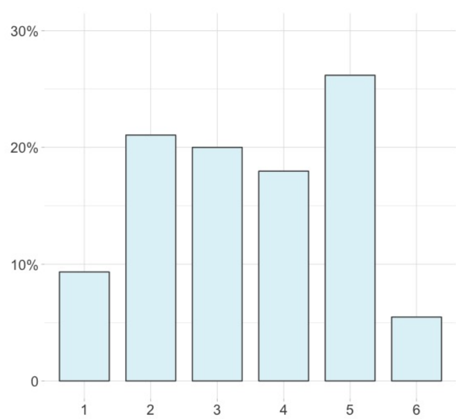


Figure 3: Knowledge complexity

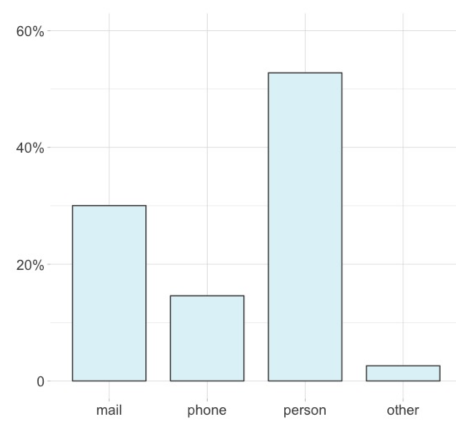


Figure 4: Mode of communication

Figure 3 depicts the distribution of the categorical variable *Knowledge*. Simple instructions (category 1) and step-by-step follow-up (category 6) were cited by less than 10% of respondents each. The most popular category 5 - analysis of the specific problem and definition of how to deal with it - resulted in 30% of citations. Half of the requests concentrate on the first three categories dealing with simple instructions, information which should be codified and easily retrievable through organizational wikis.

Figure 4 depicts the type of communication channel used. Half of respondents indicated a preference for personal meetings as the way to contact colleagues, while the usage of email (30%) prevailed over the telephone (15%).

Tables 1 and 2 summarizes the descriptive statistics for the variables.

Table 1: Descriptive statistics: categorical variables.

Variable	**Levels**	**n**	%	$\sum$%
Efficiency	1	35	2.4	2.4
	2	40	2.8	5.2
	3	86	6.0	11.2
	4	173	12.0	23.2
	5	389	27.1	50.3
	6	714	49.7	100.0
	all	1437	100.0	
Knowledge	1	141	9.8	9.8
	2	301	20.9	30.8
	3	286	19.9	50.7
	4	257	17.9	68.5
	5	374	26.0	94.6
	6	78	5.4	100.0
	all	1437	100.0	
Channel	1	436	30.3	30.3
	2	208	14.5	44.8
	3	756	52.6	97.4
	4	37	2.6	100.0
	all	1437	100.0	
Gender	f	122	40.3	40.3
	m	181	59.7	100.0
	all	303	100.0	

Table 2: Descriptive statistics: continuous variables.

Variable	**n**	**Min**	$\mathbf{q_1}$	$\widetilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	$\mathbf{q_3}$	**Max**	**s**	**IQR**	**#NA**
Age	303	27	39	48	46.4	53	68	8.8	14	0
Seniority	303	0	7	17	17.2	27	40	10.6	20	0
Load	303	1	7	15	20.1	26	202	21.1	19	0

To be able to distinguish between different levels of performance we consider $Efficiency$ as ordinal with non-interval outcomes (with minimum satisfaction coded as 1 and maximum as 6) while all the other variables are continuous ($Load$, Age, $Seniority$), categorical ($Knowledge$ and $Channel$) and dichotomous ($Gender$). Table 3 reports the results of the correlation analysis.

Table 3: Correlation table.

	efficiency	information	channel	gender	age	seniority
efficiency						
knowledge	0.381****					
channel	0.130**	0.019****				
gender	0.021	0.071	0.118			
age	−0.062****	−0.007****	0.093****	0.588****		
seniority	−0.078****	0.016****	0.028****	0.559****	0.806****	
load	−0.102****	0.036****	−0.298****	0.399****	0.127****	0.234****

*** significant level at 99%, for the Pearson correlation for numeric variables, polyserial correlation for numeric and ordinal variables and polychoric correlation for ordinal variables.

The variable reflecting employees' performance does not show strong correlations with either of the predictors. $Efficiency$ is moderately correlated with $Knowledge$, suggesting that a more complex task may result in higher efficiency. The correlation between $Gender$ and $Efficiency$ is non significant, while load, seniority, and age variables are negatively correlated with $Efficiency$. This preliminary evidence might corroborate the hypothesis that the abundance of social capital may indeed lead to the decrease in efficiency of communication and suggests that higher organizational tenure and personal experience may not be always seen as a predictor of a positive performance.

Correlation tests did not record a strong significant relationship between self-reported variables encoding the type of assistance requested and the type of communication channel used with the rest of independent variables. A moderately negative and significant correlation between channel and load variables suggest that we may find a considerable difference in the number of interactions by categories. We obtained a high positive correlation coefficient for the variables age and seniority R=0.8 [95% CI: 0.77-0.85], supporting the idea that older employees act as troves of the expertise relevant to the particular professional reality. This would also imply a significance of correspondence between higher age and seniority and higher intensity of collaborative activity (both significant, though not greatly).

3.2 The model

In order to test the impact of the various variables on the ordinal categorical variable $Efficiency$ we made use of standard ordinal logistic regression, which is widely applied as an effective method for modeling categorical outcomes with respect to its order as a function of both continuous and categorical predictors [Harrell, 2015]. Specifically we tested the following model:

$$\log(\frac{p_j^c(x)}{1 - p_j^c(x)}) = \alpha_j + \beta' x \quad \forall j = 1, ..., 6 \tag{1}$$

- $p_j^c(x) = P(Y =< j | X = x)$ represents the logit of the cumulative distributions where Y is the ordinal dependent variable measuring the efficiency of the response with value j from 1 to 6.
- $x = (x_1, ..., x_k)$ is the matrix of the k_th independent covariates.
- $\alpha_1, ..., \alpha_{r-1}$ and $\beta_1, ..., \beta_k$ the regression coefficients to be estimated.

This model is dictated by the availability of data. A natural way to measure the existence of overload would be to run a regression analysis using longitudinal data and to analyze the formation of the network in order to track whether, over time, highly requested workers exhibit a decrease in their efficiency. The data at our disposal does not allow us to carry out this exercise. Initial positive assessments and a later negative evaluation intertwine as a result of the cross-sectional nature of the survey. However, if we can still register a negative impact of *Load*, **a fortiori** we would expect to find it in a perfect scenario in which we could control for the initial positive value.

This regression method relies on cumulative distributions and fits parameters for each association to estimate a general trend across the ordinal values of the dependent variable, retaining information on the rank ordering [Agresti, 2002].

One the most fundamental assumptions of the ordinal logistic regression is the proportional odds assumptions, which require that different categories of the variable have the same effect on the outputted odds. Following Harrell [2015], a visual estimation of the linear predicted values of the given variables calculated with relaxed parallel slope assumption suggested that the variable $Knowledge$ violates the proportional odds assumption. The Brant test [Brant, 1990] shows that the effect of the variable is not constant across separate binary models, which supports the idea that the proportional odds assumption is violated. As a solution we group the 6 categories of the variable $Knowledge$ in two categories which can group the possible answers in "Simple Knowledge" and

"Complex Knowledge". The former groups knowledge transfer, which requires simple instruction or some more complex explanation of simple instructions,while the "Complex Knowledge Transfer" groups categories which involve practical demonstration with exercise and experiments, analysis of the specific problem or even a step-by-step follow up and a verification of the results.

Repeated testing with the binary version of the variable did not yield significant results [4], indicating satisfied proportional odds assumptions.

3.3 Results

The results of the model selection are reported in the Table 4 in which we report coefficients, p-values and standard errors[5]. As described above, the dependent variable $Efficiency$ is expressed as categorical and ranging from lower to higher efficiency. Thus, a positive and significant value of the estimate indicates that when keeping all the independent variables constant, with one unit increase in a continuous covariate, the odds of moving to a higher level of efficiency increases. An easier interpretation of the result in figure 5 is given by the odds-ratio. The odds ratio gives the factor of increase in the probability of moving to a higher category of efficiency when a continuous variable increases by one unit or a categorical variable change from the reference category. For instance, if the *Load* increases by one unit only, the odds of moving to a higher category of $Efficiency$ decreases by 1%.

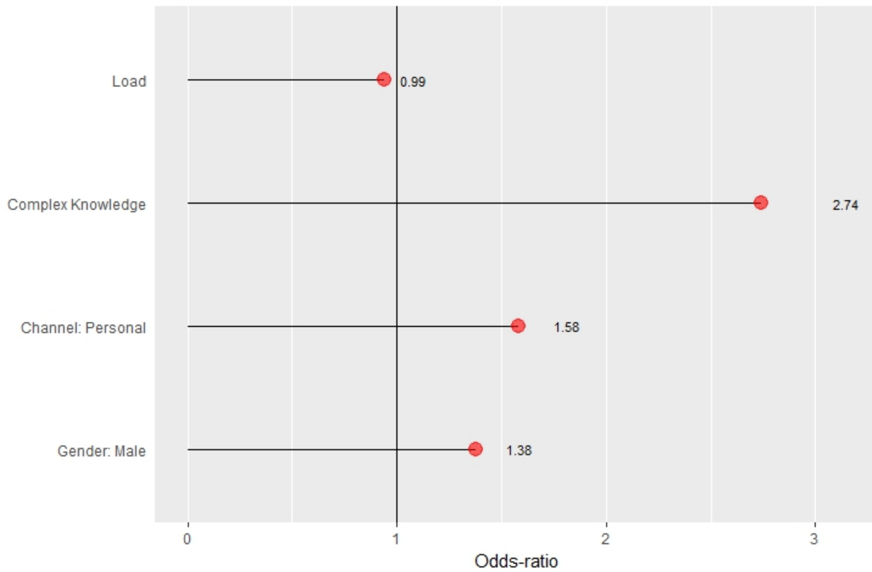
[4] $\chi^2 = 39.75$ ($p > .31$)

[5] We use the software R and the package MASS [Venables and Ripley, 2002]

Table 4: Regression Results

	Dependent variable:			
	Efficiency			
	(1)	(2)	(3)	(4)
Load	-0.004^{***}	-0.004^{***}	-0.003^{***}	-0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Knowledge (complex)		0.960^{***}	0.967^{***}	0.971^{***}
		(0.103)	(0.103)	(0.104)
Channel(Phone)			-0.474	-0.469
			(0.296)	(0.297)
Channel(Person)			0.444^{***}	0.461^{***}
			(0.121)	(0.123)
Channel(Other)			0.032	0.019
			(0.155)	(0.156)
Seniority				-0.009
				(0.010)
Age				-0.011
				(0.012)
Sex(M)				0.297^{*}
				(0.161)
AIC	3706.1	3618.5	3602.8	3601.3
Observations	1,425	1,425	1,425	1,425
Lipsitz test				p-value = 0.053
Pulkstenis-Robinson test				p-value = 0.56
Hosmer-Lemeshow test				p-value = 0.14 [6]
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 5: Odds-ratio for significant effects in Model 5



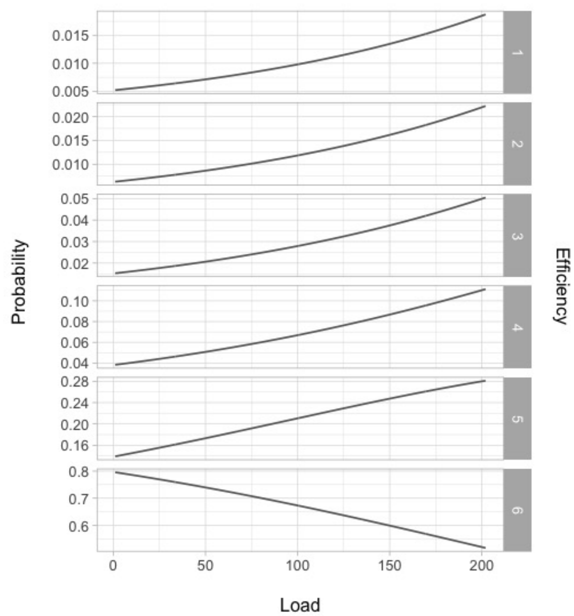
An additional way of interpreting the result is to plot the predicted probability of each class of $Efficiency$ for an increasing $Load$ as in figure 6. As expected for high level of $Load$ the predicted probability of high $Efficiency$ is much larger than for low values since employees contact colleagues who are likely to satisfy their request. However, it is also clear that with an increasing load the probability of scoring the highest level in $Efficiency$ dramatically drops and the probability is absorbed partly by all other categories.

This evidence supports the idea that increased burden of collaborative behavior by employees will eventually result in the decrease of the efficiency of collaboration.

As previously mentioned, the variable $Knowledge$ has been coded into a binary variable to satisfy OLR assumptions. Positive and significant estimates indicate that moving simple type communication to more complex ones, such as "Step-by-step follow-up of the specific activity and verifying results" increases the odds of improving the $Efficiency$ by a factor of 2.74. The effect of the knowledge variable on the efficiency remains strong when controlling for the type of communication channel used. Data shows no significant effect on $Efficiency$ as the choice for the conversation shifts from email (reference category) to telephone. However, face-to-face collaboration had a significant positive impact on the dependent variable indicating that the odds of performing better are 1.71 higher.

Collaborative activity, knowledge complexity, and face-to-face interactions remain significant predictors of positive performance when personal characteristics are controlled for. The odds of an high evaluation of efficiency is 1.38 times larger when choosing a male over a female colleague.

Figure 6: Predicted probabilities.



3.4 Discussion

As per the result of the literature analysis on the possible adverse effects of extensive collaboration, we conclude that collaboration overload poses a risk to individual performance and that organizational environment factors can partly predict employees' performance. We applied a regression analysis to identify the factors that influence collaboration efficiency among employees. The most important result of our model is the significant and negative influence of collaborative load on the efficiency of processing requests by colleagues. By estimating the predicted probabilities for increasing collaborative load, we show a clear negative impact on the highest levels of employees' efficiency in expertise sharing.

Efficiency is positively associated with the nature of information complexity, employees gain high levels of satisfaction when they process requests that require more intense interactions, such as problem-solving or step-by-step follow-ups. Similarly, the probability of delivering higher level efficiency grows when the interaction happens in person rather than through email. Finally, neither organizational tenure nor the personal experience expressed in years is a significant factor, suggesting that the informal network is independent from the formal network of an organization's hierarchy.

Undertaking research in the organizational setting of a private company poses a series of restrictions, privacy legislation being the most challenging for academic research. Limiting constraints on the data about employees' personal characteristics and a high cost of mining the information through observation and survey contribute to the poor quality of existing empirical research.

We thus contribute to the research on information complexity and its requirements for higher cognitive capacity and time. Previous studies have suggested a link between the type of knowledge and the means and outcomes of its transfer from one actor to another [Schneider, 1987, Eppler and Mengis, 2004]. This research provides sound empirical evidence that, with the rising complexity of knowledge, the intensity of interactions between actors will increase and thus the efficiency of collaboration will also increase.

This analysis adds a new perspective on the efficiency of the methods for information transfer. There was no significant impact of the age of the actors, nor seniority. Considerable low levels of telephone usage compared to electronic mail and face-to-face conversation reveal employees' vision of an effective communication technology. As expected, face-to-face communication delivers greater odds for providing higher productivity.

The results of the analysis suffer from a possible endogeneity because the formation of the observed networks depends very likely on the past and expected efficiency of the interaction. However, as discussed before, it is fair to assume that the negative effect of the collaborative load occurs later on in the network formation when it becomes excessive and creates time and cognitive constraints. If results are biased, they are underestimating the negative effect of load on efficiency.

Managers are encouraged to incorporate the role of intraorganizational communications as a firm's strategic resource, with the relevant positive and negative externalities it may bring. Organizational decision-makers should embrace the importance of recognizing the regularities of social structures in order to understand the effect that a single actor behavior has on the welfare of other organizational players. Managerial interventions should focus on two types of behavior change - directed on key actors and information seekers.

As the developed model suggests, the primary emphasis should be placed on identifying the key expertise holders among employees. Adequate coaching and mentoring programs facilitate the diffusion of knowledge possessed by stars. Quantifying collaborative load that organizational actors encounter in their everyday activity provides a benchmark for the identification of active collaborators who should be prioritized when developing retention and promotion plans.

Development of knowledge repositories provides organizational leverage to control over the stream of expertise sharing requests. Such organizational wikis summarise information that has a lower level of ambiguity and complexity, thus can be easily codified and retrieved. Another intervention to equalize collaboration should be aimed at strategical staff placement. Organizational space arrangement should favor proximity of interdependent employees to allow for face-to-face collaborations, which increase the probability of delivering higher satisfaction of information request.

4 Conclusions and Limitations

This paper investigated the existence of collaborative overload in the workplace and gave an account of possible factors contributing to social capital failure. By modeling the social structure of collaborations in the professional service firm, we reproduce a network of individual incentives that translate into network outcomes. Consequently, we identify top performers that take upon themselves the majority of peer assistance requests and estimate the probability of an efficient outcome of the collaboration between two colleagues under specific task constraints.

We thus conclude that active involvement in collaborative activity can lead to over-embeddedness of top performers reducing the probability of the establishment of efficient knowledge sharing practices. In line with the previous research [Eppler and Mengis, 2004, Oldroyd and Morris, 2012, Cross et al., 2002, Abbasi et al., 2014], the results of this study support the idea about the presence of few focal actors, who possesses greater visibility in the organizational network, acting as a reference point for the majority of colleagues. Secondly, with the growth of collaborative load, we observe a negative impact on the performance of those actors. The effect of collaborative load remains statistically significant even after controlling for communication mode, organizational context, and personal factors.

Finally, a number of potential limitations encountered during the design of this study should be mentioned. Despite various determinants used for the evaluation of social capital, there is a general consensus on the low validity of delivered results [Falk and Harrison, 1998]. Performance is the most complicated measure in the field of communication research, obscuring obtained results and possible interventions. The data used for defining our variable of interest, employees' productivity, was self-reported, and is therefore a potential source of response bias. While the survey methodology allows the type of communication appropriate for the research to be isolated, constructed network of organizational collaborations has a static nature, which restricts the predictive power of the produced model. Also, our decision to address a particular reality of the professional services industry limits generalizability of the obtained result. Contextual factors related to the specific organizational setting may produce different results on employees' performance. Future empirical tests must be performed to evaluate the efficiency of the suggested approach in different contexts. Proposed model could profit from the integration of performance indicators that will complement colleagues' satisfaction rate with the factual results. Analyzing communications in a longitudinal perspective or running a control survey with a time interval may deliver more powerful predictions.

All in all, these findings contribute to both academic and business research. We add to a growing body of literature on risks posed by the excess of information requests placed especially on a few key individuals, affecting their performance, and ultimately damaging the proper function of the organization. Results do not suggest the reduction of intraorganizational communications, but they rather signal inefficiency in the knowledge and teamwork management. As managers stress the necessity and create more possibilities for teamwork, the pressure on key employees is rising. Application of this approach to real organizational data enriches human resource departments with the knowledge of intraorganizational communication patterns and activity of knowledge workers, potentially at risk of being burdened by overload. We suggest that the results obtained should be transformed into performance indicators to be used when evaluating the efficiency of collaborative patterns, identification of attrition sources, and, finally, should be seen as a premise for the improvement of reward mechanisms. An organization that fails to embrace the necessity for communication patterns analysis runs the risk of creating a disconnect in employees' engagement, thus contributing to the rise of barriers to inclusiveness. Though the value created by employees' interactions is technically hard to measure, it should not be seen as practically impossible.

5 Appendices

Survey

1. **Indicate five colleagues to whom you came for help and from whom you received the necessary knowledge or information required to perform a certain task in the last six months?**

   (a) Use the scale from 1 to 6 to evaluate how satisfying colleagues performance has been (1=minimum and 6 = maximum).

   (b) Use the scale from 1 to 4 to indicate the frequency of interactions with the colleague (1=minimum and 4 = maximum).

2. **Indicate the type of information that describes your request:**

   1. Simple instructions.
   2. Instructions with more than one interaction for more clarifications.
   3. More complex explanation of simple instructions, with greater interaction.
   4. Practical demonstration with exercises and experiments on cases.
   5. Analysis of the specific problem and definition of how to deal with it.
   6. Step-by-step follow-up of the specific activity and verifying results.

3. **Indicate the type of communication channel that was used for contacting a colleague:**

   1. Email
   2. Telephone
   3. Personal meeting
   4. Other

References

A. Abbasi, R. T. Wigand, and L. Hossain. Measuring social capital through network analysis and its influence on individual performance. *Library & Information Science Research*, 36(1):66–73, 2014.

P. S. Adler and S. W. Kwon. Social Capital: Prospects for a New Concept. *The Academy of Management Review*, 27(1):17–40, 2002.

F. Agneessens and R. Wittek. Where do intra-organizational advice relations come from? The role of informal status and social capital in social exchange. *Social Networks*, 34(3):333–345, 2012.

A. Agresti. *Categorical Data Analysis*, volume 45. Jonh & Wily Sons., New York, 2002.

D. Allen. Information overload: context and causes. *The New Review of Information Behaviour Research*, 4(1):31–44, 2003.

R. Andrews. Organizational social capital, structure and performance. *Human Relations*, 63(5): 583–608, 2010.

S. Ansari, K. Munir, and T. Gregg. Impact at the 'Bottom of the Pyramid': The role of social capital in capability development and community empowerment. *Journal of Management Studies*, 49(4):813–842, 2012.

S. Behrendt, A. Richter, and K. Riemer. Conceptualisation of Digital Traces for the Identification of Informal Networks in Enterprise Social Networks. *25th Australasian Conference on Information Systems*, pages 1–10, 2014.

S. P. Borgatti and R. Cross. A relational view of information seeking and learning in social networks. *Management Science*, 49(4):432–445, 2003.

R. Brant. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics, 1171-1178. *Biometrics*, pages 1171–1178, 1990.

R. Burt. *Structural holes.*, volume 100. Harvard University Press, Cambridge, MA, 1992.

J. S. Coleman. Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94:S95–S120, 1988.

R. Cross and P. Gray. Where Has the Time Gone? *California Management Review*, 56(1):1–17, 2013.

R. Cross, S. P. Borgatti, and A. Parker. Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, 44(2):25–47, 2002.

R. Cross, R. Rebele, and A. Grant. Collaborative overload. *Harvard Business Review*, 94(1):16, 2016.

A. F. de Toni and F. Nonino. The key roles in the informal organization: a network analysis perspective. *The Learning Organization*, 17(1):86–103, 2010.

T. Dewett and G. R. Jones. Dewett, T. and Jones, G.R., 2001. The role of information technology in the organization: a review, model, and assessment. *Journal of management*, 27(3):313–346, 2001.

C. Di Ciccio, A. Marrella, and A. Russo. Knowledge-intensive Processes: An overview of contemporary approaches. *CEUR Workshop Proceedings*, 861:33–47, 2012.

K. Dooley. Organizational complexity. *International encyclopedia of business and management*, pages 5013–5022, 2002.

A. L. Downs, C. W., Clampitt, P. G., Pfeiffer. Communication and organizational outcomes. In *Handbook of organizational communication*, pages 171–211. 1988.

L. Empson. *Introduction: Knowledge management in professional service firms*. 2001.

M. J. Eppler and J. Mengis. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society: An International Journal*, 20(5):325–344, 2004.

M. W. Fagerland and D. W. Hosmer. Tests for goodness of fit in ordinal logistic regression models. *Journal of Statistical Computation and Simulation*, 86(17):3398–3418, 2016.

I. Falk and L. Harrison. Indicators of social capital: Social capital as the product of local interactive learning processes. 1998.

L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

F. Fukuyama. Social capital and development: The coming agenda. *SAIS Review*, 22(1):23–37, 2002.

G. Goldhaber and G. Barnett. *Handbook of Organizational Communication*. Buffalo, New York, 1988.

M. Granovetter. The strength of weak ties: A network theory revisited. *Sociology Theory*, pages 201–233, 1983.

F. E. Harrell. Ordinal Logistic Regression. In *Regression Modeling Strategies*, pages 331–325. Springer, 2015.

G. P. Huber. A Theory of the Effects of Advanced Information Technologies on Organizational Design, Intelligence, and Decision Making. *Academy of Management Review*, 15(1):47–71, 1990.

E. Ibarra. Network Centrality, Power, and Innovation Involvement : Determinants Of Technical and Administrative Roles. *The Academy of Management Journal*, 36(3):471–501, 1993.

M. O. Jackson. *Social and economic networks.* Princeton university press, 2010.

E. Jones, L. Chonko, D. Rangarajan, and J. Roberts. The role of overload on job attitudes, turnover intentions, and salesperson performance. *Journal of Business Research*, 60(7):663–671, 2007.

R. Kaše, J. Paauwe, and N. Zupan. HR practices, interpersonal relations, and intrafirm knowledge transfer in knowledge-intensive firms: A social network perspective. *Human Resource Management*, 48(4):615–639, 2009.

D. Krachardt and J. Hanson. Informal networks: the company behind the chart. *Harvard Business Review*, 71(4):104–111, 1993.

D. Krackhardt. Assessing the Political Landscape: Structure, Cognition, and Power in Organizations. *Administrative Science Quarterly*, 35(2):342–369, 1990.

M. Lazarova and S. Taylor. Boundaryless careers, social capital, and knowledge management: Implications for organizational performance. *Journal of Organizational Behavior*, 30(1):119–139, 2009.

R. Lee and O. Jones. Networks, communication and learning during business start-up: The creation of cognitive social capital. *International Small Business Journal*, 26(5):559–594, 2008.

P. R. Monge and N. S. Contractor. *Theories of communication networks.* Oxford University Press, USA., 2003.

E. Muller and D. Doloreux. What we should know about knowledge-intensive business services. *Technology in Society*, 31(1):64–72, 2009. ISSN 0160791X. doi: 10.1016/j.techsoc.2008.10.001. URL http://dx.doi.org/10.1016/j.techsoc.2008.10.001.

J. B. Oldroyd and S. S. Morris. Catching falling stars: A human resource response to social capital's detrimental effect of information overload on star employees. *Academy of Management Review*, 37(3):396–418, 2012.

P. Palmgreen. Uses and gratifications: A theoretical perspective. *Annals of the International Communication Association*, 8(1):20–55, 1984.

M. Polanyi. *The tacit dimension.* Garden City, NY, 1967.

W. C. Redding and P. K. Tompkins. Organizational communication: Past and present tenses. In *Handbook of organizational communication*, pages 5–33. Norwood, NJ: Ablex Publishing, 1988.

S. C. Schneider. Information overload: Causes and consequences. *Human Systems Management*, 7 (2):143–153, 1987.

C. Speier, J. S. Valacich, and I. Vessey. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2):337–360, 1999.

W. Tasi and W. Tsai. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit, innovation and performance. *Academy of Management Journal*, 44(5):996–1004, 2001.

D. Tjosvold and Y. Tsao. Productive Organizational Collaboration: The Role of Values and Cooperation. *Journal of Organizational Behavior*, 10(2):189–195, 1989.

W. Tsai and S. Ghoshal. Social Capital and Value Creation: The Role of Intrafirm Networks. *The Academy of Management Journal*, 41(4):464–476, 1998.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S.* Springer, New York, fourth edition, 2002.

O. A. Wiio. *Contingencies of organizational communication: studies in organization and organizational communication.* Helsinki School of Economics, Helsinki, 1978.

B. Yamkovenko and J. P. Hatala. Methods for Analysis of Social Networks Data in HRD Research. *Advances in Developing Human Resources*, 17(1):40–56, 2015.

Emergence of the knowledge sharing practices.

Anna Velyka [1,2]

[1] Despina, Big Data Lab, Department of Economics and Statistics "Cognetti de Martiis"

[3] Aizoon

Abstract

This paper examines the underlying social processes influencing the emergence of organizational knowledge transfers. High degree of tacitness inherent to the personal knowledge of employees and the lack of commonly adopted methodology to measure it create obstacles for the inclusion of knowledge transfer practices in performance evaluation. The data about knowledge transfers originate from the survey conducted in the professional service industry. This paper uses an exponential random graph modeling for the analysis of the structure, sharing patterns and significant factors driving the emergence of informal networks.

Understanding the mechanisms leading to the emergence of knowledge transfers allows to identify organizational knowledge hubs and reveal effective work arrangements. This research confirms previous findings indicating that knowledge transfers with prevailing codified elements are more likely to happen within an organizational unit boundary and among actors communicating daily. On the contrary, highly tacit knowledge transfers are less likely to arise among actors located physically close to each other (the same room). This result presents the argument in favor of organizing specialized learning groups and community-of-practices. Classification of knowledge used for this analysis identifies a group of transfers that possess both implicit and explicit components. The mechanisms leading to the emergence of interaction in this group of knowledge require further investigation to reveal crucial differences from strictly codified or tacit networks.

Keywords: knowledge transfers; organizational interactions; exponential random graph model;

1 Introduction

All knowledge has a tacit component and the degree of tacitness varies. Polanyi [1967] attributes the level of tacitness to a "personal knowing", being difficult to express and to convey. As such knowledge tacitness is a function of transfer capacity rather than the nature of the information. A division of knowledge into two categories (tacit and codified) inherent to the knowledge management literature is erroneous and may lead to poor conclusions [Grant, 2007].

The efforts to categorize knowledge by tacit component date back to the works of Nonaka [1991], Nonaka and Hirotaka [1995], Nonaka and Konno [1998]. Authors stressed the defining role of working space and social arrangements to enable the conversion of tacit knowledge into explicit. Knowledge embodied in countless formal and informal interactions is a critical corporate asset. Uncodified and subjective nature intrinsic to knowledge may hamper absorptive capacity. Knowledge corporatization implies the process of creating collaboration practices and trust between individuals and groups to facilitate the transformation or personal information into organizational.

Organizational knowledge is composed of individual intellectual capital coming together within organizational boundaries. Personal knowledge becomes embedded in the organizational structure through codification in documents, procedures, norms, emergent operational routines and practices [Davenport and Prusak, 2000]. The growing importance of collaborative arrangements favors knowledge exchange between organizational actors. Examination of intraorganizational communications evolved into a benchmarking framework for gaining insights into collaborative practices[Cross et al., 2003, Monge and Contractor, 2003, Johnson et al., 2012], and knowledge transfers [Grippa et al., 2006, Zappa and Robins, 2016]. Informal organizational networks provide a quantitative overview of knowledge sharing practices [Cross and Parker, 2004].

The body of literature discussing possible approaches to classification of knowledge is growing [Christensen, 2007, Omotayo, 2015, Odero and Otuya, 2017]. Previous research focused on two main aspects: the origin of knowledge (individual, organizational routines formal guidelines) and the form it acquires (organizational, tacit, explicit). A rigid approach to categorization may fail to account for the instrumentality of knowledge [Christensen, 2007, Davenport et al., 1998]. Christensen [2007] argues that the role of knowledge is defined not only through the transfer of the best practices, but also through bridging interdependencies.

Interdependencies emerging on the grounds of individual intrinsic motivation provide insights into advice seeking, expertise sharing, and knowledge transfer [Krachardt and Hanson, 1993, Zappa and Robins, 2016]. Focusing the analysis on the type of knowledge that is shared, expands managers understanding of possible interventions in the work arrangements. This paper presents an inquiry into organizational know-how transfers. I use exponential random graph modeling to measure the impact of node, edge and network level predictors on knowledge sharing practices in the organization. This methodology was suggested by Monge and Contractor [2003] for testing the significance of various multilevel attributes on the probability of tie emergence and later applied for investigation of the factors that shape the network structure [Quintane and Kleinbaum, 2011, Johnson et al., 2012, Robins, 2013, Zappa and Robins, 2016, Pilny and Atouba, 2018].

The study relies on the data obtained through the questionnaire conducted among 514 employees of a medium size technology consultancy firm in Italy. I use collected data to reproduce the networks of knowledge exchange implying different degree of tacitness. The primary hypothesis verified in this paper is the existence of a significant relationship between the structure of socialization and the degree of knowledge tacitness. The three sub-hypothesis tested in this research uncover mechanisms influencing the knowledge exchange practices.

2 Research Framework

Decision-making process in the organizational setting is shaped by the personal knowledge. Individuals gain this knowledge through learning, experiencing and interacting with peers. Polanyi [1962] argues that knowledge acquired in this way constitutes a high tacit component, that eventually acquires an explicit nature once being shared [Polanyi, 1962]. This paper relies on the definition suggested by Christensen [2007] according to whom knowledge sharing is a process of "identifying existing and accessible knowledge, in order to transfer and apply this knowledge to solve specific tasks better, faster and cheaper than they would otherwise have been solved".

With regards to organizational context, knowledge tacitness is regarded as a continuum. Once an individual becomes a part of an organization, his or her knowledge becomes a shared asset. Managers strive to create working arrangements that would increase an absorptive capacity of the knowledge resource to transform it into organizational outcomes. Nonaka and Hirotaka [1995] suggests that the articulation of tacit and explicit component should follow a spiral cycle: through application implicit knowledge acquires explicit nature to be subsequently absorbed as implicit personal knowledge. This transformation cycle favors corporatization of personal insight and provides indications for knowledge management practices.

A more detailed classification of knowledge specific to organizational context should fit tacit-explicit continuum [Blankenship and Ruona, 2009]. According to a knowledge taxonomy developed by Alavi and Leidner [2001] tacit knowledge is rooted in personal actions, experiences, and involvement in a specific context. It gains an explicit nature when articulated and generalized. When a tacit component of the knowledge transfer does not lend itself to codification, it is absorbed through the application. Otherwise, if a tacit component is formalized, knowledge is passed on through communication processes [Grant, 1996]. The two dimensions are mutually dependent rather than dichotomous [Alavi and Leidner, 2001]. As such the distinction between tacit and explicit knowledge becomes manifest in the process of transfer between employees across space, and time [Grant, 1996].

Social relationships reflected through formal and informal dimensions play a crucial role in knowledge sharing [Granovetter, 1973, Burt, 1992, Borgatti and Cross, 2003] providing mechanism for individual expertise transfer. Often, formally registered work processes overshadow the role of informal arrangements, depriving them of being understood or supported by the organizational management. Cross et al. [2002] argued about the importance of performance implications preserved in effectively functioning informal networks. Understanding of the working mechanism of informal networks becomes possible through the structural analysis of individual, group, and network level factors leading to its emergence. The professional service organization observed in this paper is an example of a tiered system that gives rise to distinct groups of interdependencies within which employees collaborate to perform daily tasks and encounter a decision-making process.

In carrying out day-to-day functional duties an individual relies on the knowledge formed as the combination of formal education and individual experience. Christensen [2007] defines this dimension as **professional knowledge** with a high concentration of tacit component. An exchange of this type of expertise plays a critical role in the operational activity, yet its contribution to the organizational outcome is indirect and is not traditionally measured. While relying heavily on personal intuition for decision-making individuals are also guided by the series of established organizational routines, rules, and procedures [Blankenship and Ruona, 2009]. Such arrangements constitute **coordinating knowledge** that facilitates the application of professional expertise. Given its prescribed nature coordinating knowledge is expected to be well codified and highly explicit [Christensen, 2007, Blankenship and Ruona, 2009]. **Object-oriented knowledge** lies between the tacit and explicit ends of presented progression. It is directly connected to individual expertise while being applied explicitly to organizational practice. In this manner, high reliance on tacitness of personal knowledge complements the explicitness of directions for the application outcome Christensen [2007].

Corman and Scott [1994] suggested illustrating the network of operational interdependencies through studying individual reflection on the knowledge transfer processes obtained through the survey. The difference in the nature of knowledge dictates the choice of the transfer channels [Polanyi, 1967], absorptive capacity [Tasi and Tsai, 2001], and learning process [Zappa and Robins, 2016]. I base this research on the assumption that the nature of knowledge defines the structural difference in the networks, and thus the requirements for effective collaboration mechanisms.

Hierarchical order and physical arrangements create differential opportunities for employees [Ibarra, 1993, Kleinbaum et al., 2013]. Functional departmentalization results in accumulation of specific operational knowledge that underlies the ability to produce distinct products and services [Keith, 1998]. Departments are harbors inside which employees perform their daily activity in cooperation with homogeneous colleagues, generating specific knowledge [Zappa and Robins, 2016]. They provide benefits of physical proximity, fast and frequent interaction. An organization defined by the structure operates through processes [Thompson, 1967]. Knowledge transfer is a process that is bridging operational interdependencies, through a combination of different forms of knowledge [Thompson, 1967, Christensen, 2007]. When information available within the department borders is not enough for effective decision-making, an individual may choose to reach out to heterogeneous colleagues in other departments to retrieve necessary information. The concentration of specific expertise inside a particular unit attracts peers seeking information and brings about cross-boundary interactions.

Another factor facilitating interdepartmental knowledge transfers is physical proximity. Frequent interactions facilitated by the proximity arrangements increases the probability of serendipitous interactions [Allen, 1977, Krackhardt, 1994] leading to the emergence of informal links creating a possibility to access the knowledge specific to heterogeneous actors [Soda and Zaheer, 2012, Kleinbaum et al., 2013, Zappa and Robins, 2016]. Individuals discover each other and based on the information about access and costs of potential resources make decisions about future interactions [Borgatti and Cross, 2003]. Thus, locations at the workplace dictate the structure of informal interdependencies. The requirements on the nature of collaboration arrangement between two individuals sharing the knowledge will differ with regards to a tacit-explicit continuum of the knowledge. Knowledge with prevailing tacit component will require more personal interaction. Well codified knowledge regarding operational activity is expected to be concentrated in the administrative departments.

Hypothesis 1: The probability of observing a knowledge transfer between employees is conditioned by formal organizational affiliation and physical proximity.

A decision to create a knowledge transfer tie with homogeneous peers reflects an individual belief about possible benefits preserved in the majority of which he or she is a part. Established operational interdependencies contribute to the formation of employees beliefs about colleagues' expertise. Individuals are more likely to seek knowledge from the colleagues whose areas of expertise are known to them [Borgatti and Cross, 2003]. The logs of email communications is an approximation of how hierarchical structure and individual choices come together in organizational context [Krachardt and Hanson, 1993, Kleinbaum et al., 2013, Zappa and Robins, 2016].

Digital record-keeping and retention of information through the store-and-forward model of electronic mailing services facilitates information flow. A higher level of credibility [Moenaert and Souder, 1990] of the written communication channel made it a primary tool for diffusion of formal instructions and externalization of codified knowledge generated within different groups [Grippa et al., 2006]. This channel can be a less efficient choice for the dissemination of unstructured knowledge, that imposes higher requirements on comprehensibility, or requires a greater degree of trust [Moenaert and Souder, 1990, Grippa et al., 2006]. Such type of knowledge transfer necessitates face-to-face communication manifesting itself in the network of prearranged interdependencies. As such stronger working ties can reflect the structure of knowledge transfer depending on the nature of knowledge shared.

Hypothesis 2: The probability of observing a knowledge transfer between employees is conditioned by the higher frequency of work-flow links.

Professional service industries rely heavily on the human capital resource. Social connections play a crucial role in the sharing of knowledge and providing inputs to problem-solving. Uncovering the structure of informal networks in which ties between actors represent information-sharing interactions provides insights into how the decision-making process is made. Statistical network modeling estimates the probability of the tie emergence under the influence of various dyadic dependent effects. One of the most frequently expected configurations dictating tie emergence is reciprocity. This statistic estimates whether the mutuality of knowledge transfer between employees happens more commonly than by chance.

Each node exercises power over the network structure insofar as it can create new or modify existing links [Knoke and Yang, 2008]. Nodes position in the knowledge flow network defines its ability to dictate the evolution of the system [Cross et al., 2002]. Employees accumulating the highest number of incoming ties are the knowledge hubs that enjoy the higher level of control over decision making and organizational structure connectivity [Cross et al., 2002]. In-degree distribution estimates network popularity spread. These statistics evaluate if the network tends to centralize around fewer nodes of a higher degree. The out-degree distribution estimates activity spread [Hunter, 2007]. These endogenous dependency structures facilitate model convergence and ensure satisfying goodness of fit [Hunter, 2007].

Edge-wise shared partner statistics allows evaluating the mediator effect [Snijders et al., 2006]. By estimating statistical power exercised by the presence of none, one, or two shared partners on the emergence of knowledge sharing tie the model suggests if knowledge transfers are more likely to form at random, or as a result of friendship and cooperation arrangements. The presence of the tacit component may impose higher demands on the level of trust between individuals. Trust may be built as the result of the previous interaction, or higher visibility, feedback from the colleagues. Exogenous covariates estimating the network structure are expected to differ among analyzed knowledge transfer networks under the influence of the extent of tacitness inherent to the knowledge nature. Identifying exogenous effects influencing informal ties emergence between employees provides indications for organizational coordination. The stronger statistical power of organizational homophily and individual attributes in comparison with reciprocity and transitivity effects uncovers mechanisms for ensuring more efficient diffusion of knowledge.

Hypothesis 3: The probability of observing a knowledge transfer between employees is conditioned by the structural attributes of the network.

A multitheoretical multilevel perspective suggested by Monge and Contractor [2003] provides a theoretical framework for analyses of the network evolution, accounting for various dydic dependent and independent effects. Exponential random graph modeling (ERGM) is gaining popularity in the field of social sciences as a statistical modeling tool for the estimation of the myriad of mechanisms shaping social ties formation [Monge and Contractor, 2003, Pilny and Atouba, 2018]. This paper applies exponential random graph modeling to empirically test the hypothesis about the decisive role the nature of knowledge plays in the emergence of intraorganizational social structures.

3 Design and Methodology

3.1 Methodology

Methodological framework applied for the study of knowledge transfers originates from the theories of communication and coordinating [Krachardt and Hanson, 1993, Monge and Contractor, 2003, Colfer and Baldwin, 2016]. General linear modeling limits the explanation of the emergence of network structure to either individual or collective influence [Robins, 2013]. Multitheoretical multilevel framework eliminates this limitation and offers a broader understanding of the fundamental processes underlying the evolution of organizational interactions enabling simultaneous analysis of individual, dyadic and global factors [Monge and Contractor, 2003].

Exponential Random Graph Modelling has been widely used in the field of social sciences to investigate the emergence of ties [Monge and Contractor, 2003, Johnson et al., 2012, Robins, 2013, Pilny and Atouba, 2018]. It is a flexible statistical method that broadens the amount of variance explained by different factors through a series of simulation of the random processes determining the possibility of a tie formation [Monge and Contractor, 2003]. The theory underlying ERGM derives from the assumption that actors' propensity to form ties is affected by a myriad of intrinsic microstructures. Thus, the focus of analysis shifts from a node to a dyad that emerges under the influence of various endogenous and exogenous covariates. A coefficient calculated for every predictive variable reveals the nature and the significance of the odds of a tie formation conditional on existing ties.

In this paper, ERGM is applied to model the probability of knowledge transfer between employees as a function of individual characteristics and network structure. A set of nodes (representing employees) and directed links (representing knowledge transfers) depict an emergent network of social relations. For the matter of this analysis exponential random graph modeling provides an optimal statistical framework to account for dependence in tie formation as knowledge sharing practices between two nodes *i* and *j* may be influenced not only by individual characteristics of two nodes, but also of other nodes in the network and the structure of the network itself. I use the following model to calculate the general probability distribution of a graph with a given number of nodes.

$$Pr(Y = y) = (1/k)exp\left\{\sum_A \eta_A g_A(y)\right\}$$

- $\sum_A$ the summation over all subset of variables, defining a model configuration.
- η_A parameter corresponding to each fitted configuration A (is non-zero only if all pairs of variables in A are assumed to be conditionally dependent);
- $g_A(y)$ the number of configurations A occurring in the network y;
  $gA(y) = 1$ configuration observed in the network y
  $gA(y) = 0$ configuration observed in the network y;
- k is a normalizing constant included to ensure a proper probability distribution. [Pattison and Robins, 2008].

The probability of observing a particular graph y arising as a result of statistics $gA(y)$ is dependent on non-zero parameters η_A of all fitted configurations A. Configurations estimated in this paper will include dyadic independent effects (individual attributes) and dyadic dependent effects (structural attributes).

Dyadic independent terms included in the model summarise homophily and multiplexity effect. Homophily effects consider both members of the dyad *i* and *j* and assess how similarity of a particular attribute influences the probability of a tie to be formed between them. Estimating homophily allows testing if the difference in organizational affiliation (specialization and proximity) and individual characteristics (gender, age, and seniority) increase the odds of a tie formation. Edge covariate is measured for each dyad and evaluates the effect of exogenous relationship on the probability of a tie presence between *i* and *j* in a given network under a condition of its existence in another network.

Knowledge transfer networks are build as directed. Thus, to estimate structural processes influencing network formation I include the following *dyadic dependent terms*: mutuality, degree distribution, and transitivity. Mutuality calculates the likelihood of a tie *i,j* to be formed under the condition of presence of a tie between *j,i*. Geometrically weighted degree distribution adds an endogenous predictor for popularity(indegree) and activity effect(outdegree). Both terms estimate the likelihood of a tie formation as a function of the degree of other nodes in the network. I am fixing decay parameters for these two configurations assist the convergence of the model [Hunter, 2007]. To evaluate transitivity and the tendency to cluster the model estimates a statistic that evaluates nodes tendency to form ties conditioned by the presence of shared partners.

Monte Carlo maximum likelihood estimation calculates parameter corresponding to a selected set of configurations and defines its significance. This algorithm tests initial coefficients of predictors on a random sample of simulated networks until converging as close as possible to the initial graph. MCMC chains of the fitted covariates are expected to follow a normal distribution. Three models fitted for professional, coordinating and object-based knowledge transfers are analyzed in R language, using ergm package [Handcock et al., 2018].

3.2 Data

Data used for this analysis come from a medium size company working in the sphere of information technology consulting. This professional activity is highly reliant on human capital knowledge. The sample consisted of 514 employees (total number of employees at the beginning of the survey).

Descriptive characteristics of organizational actors include gender, age, seniority and organizational departmentalization with no missing values. Table 1 shows summary statistics of organizational demography. Overall, women make up less than 30% of the workforce composition. 50% of employees have less than 3 years of working experience within the organization.

Organizational structure for the matter of this analysis regards two arrangements: proximity (colleagues located in the same room) and departmentalization (colleagues assigned to the same department according to professional expertise). To examine the power of proximity on the likelihood of the knowledge transfer employees are aggregated according to the physical arrangement in the office - 22 offices. Based on professional expertise employees are aggregated in 9 organizational units: sales (OU1), technical expertise (OU2,OU3,OU4), administration (OU5), human resource (OU6), staffing (OU7), and overseas subdivisions (OU8 and OU9).

The first unit responsible for sales and sales facilitation activities accounts for the highest number of employees - 331 (64% of total workforce). The demographic composition of this unit

Table 1: Descriptive statistics of the sample.

		Count	Percent
Orgnizational Unit	OU1	331	64%
	OU2	40	8%
	OU3	33	6%
	OU4	50	10%
	OU5	17	3%
	OU6	4	1%
	OU7	23	5%
	OU8	5	1%
	OU9	11	2%
Gender	Male	372	72%
	Female	142	28%
Seniority	<= 1 year	187	36%
	2-6 years	162	32%
	>= 7 yeas	165	32%
Age	21-30	101	20%
	31-40	220	43%
	41-49	147	28%
	50+	46	9%

replicates one of the general sample. The next three units in order of size are technology units that account for 123 employees (24% of total workforce). Employees of this unit engaged in engineering and application of scientific knowledge for practical purposes. These operational units have the lowest participation of women (21%) and is the youngest regarding average age (34 years) and length of service (75% of employees have the seniority of less than 5 years). (REWRITE)

Departments involved in the administrative, human resource and staffing activities (7% of total workforce) realign organizational gender balance with the majority of female workers (65%). Median seniority is 6 years. in one operational unit, represented mainly by administrative and sales personnel. The maximum seniority in both offices is 3 years.

Data obtained through the survey reflects the social structure of the knowledge transfers analyzed in this paper. Each employee was asked to identify five colleagues with whom he or she had a live informal interaction within and outside the normal work routine during the period of the last six month in regards to:

- solutions linked to existing procedures and practices (coordinating)
- support in growing job-required skills (object-based)
- transformation of new ideas into real applications (professional)

Distinctive features of the company's operational activity determine the wording of the question see Appendix 1. Throughout the paper, I will use the names specified in brackets to refer to the corresponding emergent graphs for easy reference. The survey had a high response rate: 93% of complete responses and 3% of partial responses. Partial responses were included in the analysis.

Email gained popularity as a communication tool used in an organizational setting mainly owing to its technological flexibility and time efficiency in interaction with an unlimited number of colleagues. In the given corporate context email is formally established as a principal channel for communication. The history of email logs offers a reliable reflection of the intraorganizational communication structure. While reducing the distance between employees and increasing efficiency of information transfers email data creates limitations for the empirical analysis. Mass mailings delivering information to multiple people simultaneously create noise and distort information about socially meaningful interpersonal interactions. As such extracted dataset summarising email logs for the six months preceding the day of the beginning of the survey was reduced to obtain a more accurate reflection of the knowledge transfer flows.

I remove the messages with more than four recipients. The choice of the selected threshold relies on the previous research performed in the field of organizational network analysis of email exchanges [Kossinets and Watts, 2006, Quintane and Kleinbaum, 2011]. The sample contained only email messages exchanged between employees of the given company. Information about timestamp, the subject of the email and its content is disregarded for this analysis. I construct the network by aggregating multiple links between two nodes into weighted dyads. Each dyad represents the intensity of communication between two employees (roughly the number of emails exchanged). Obtained data represents the strength of the workflow tie.

I anonymize employees identities with encrypted ID numbers, that is later used to aggregate analyzed data sets.

4 Results and Discussion

The three graphs used for the analysis contain 514 anonymized nodes. Graphs vary in the number of edges. Transfers of coordinating knowledge account for the highest amount of interactions - 924 edges, followed by object-based knowledge transfers - 669 edges, and professional knowledge transfers resulting in 378 edges. The graph of email communications accounts for 9 936 edges. Table 2 reflects network statistics for all graphs.

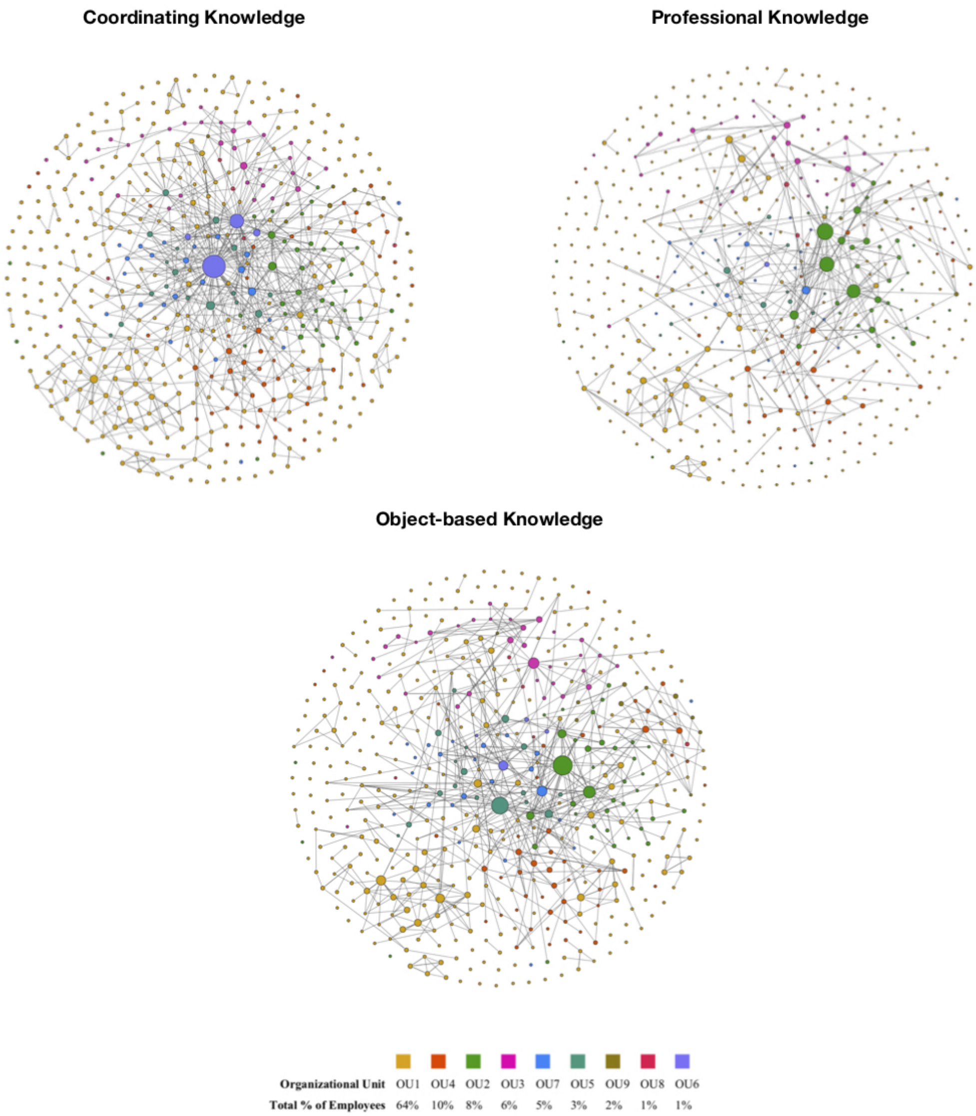
Table 2: Descriptive statistics of the networks.

	edges	density	transitivity	reciprocity
Coordinating Knowledge	924	0.0035	0.124	0.156
Object-based Knowledge	669	0.0025	0.230	0.143
Professional Knowledge	378	0.0014	0.230	0.185
Email Communications	9 936	0.0377	0.281	0.794

Due to the restriction for the number of possible citations (not more than five people) and the static nature of the networks constructed from the survey data, the density of three graphs reflecting knowledge transfers is lower than the email communication graph. Graphs constructed from self-reported data rarely exhibit a high level of transitivity as not every pair of nodes results to be connected. Transitivity estimations for email, object-based and professional graphs reports similar numbers, while it is twice as low for coordinating knowledge transfer. As expected, reciprocity in the email graph is high and low in the survey graphs signaling the presence of knowledge hubs concentrating high number of citations.

Figure 1 illustrates three graphs build from the survey data with fixed coordinates of the nodes. Indegree count defines the size of nodes, and the organizational unit determines the color. The majority of employees from the biggest functional unit (sales) lie on the periphery of the graphs showing a propensity to affiliation homophily.

Figure 1: Knowledge Transfer Networks



A graph of coordinating knowledge transfers is more sparse and has the lowest clustering coefficient (0.124). Organizational unit responsible for staffing accounts for 2% of the workforce and concentrates the most significant amount of knowledge and expertise about corporate practices (3 employees of this OU account for 12% of citations in the problem-solving network). The nodes depicting employees of technological unit constitute the core of the professional knowledge graph concentrating 12% of incoming requests. Transitivity statistic of professional knowledge graph is the same as of object-based graph (0.230), with almost twice as fewer edges.

With more equally distribute citations the graph reflecting object-based knowledge transfers renders more visible the nodes of administrative and geographically detached units. Central nodes belonging to a technological unit in the professional knowledge graph and central nodes belonging to a staffing unit in the coordinating knowledge graph concentrate a higher number of citations in the graph depicting object-based knowledge transfer. A visual analysis suggests that the graph depicting object-based knowledge ties combines structural characteristics of the other two, confirming the foundational hypothesis about the combination of tacit and explicit components.

Obtained results suggest that organizational affiliation can be a predictor for identifying knowledge hubs. The majority of top performers in the professional network belong to the technological unit; the staffing unit is a more probable trove of knowledge for an organizational praxis. However, descriptive and visual analysis is not enough to define the underlying social processes generating knowledge transfer arrangements. I use exponential random graph modeling to test the hypothesis about the significance of the tacit-explicit component effect on the structural difference between social constructs. Using information criterion approach (AIC) I construct the model predicting the likelihood of the knowledge transfer formation [Hunter et al., 2008]. Appendix B presents the final model of configurations able to reproduce the distribution of the observed knowledge transfers. To facilitate results interpretation I compose Table 3 that presents the summary of significant statistics identified by "X" for every network.

Baseline model fits a constant covariate *edges* that accounts for the number of ties and reflects the log-odds of any tie existence as independent of each other [Robins, 2013, Pilny and Atouba, 2018]. Statistical technique of estimating the effect of the number of edges is applied to define the tendency of the observed variable towards subsequently added dyadic properties [Wasserman and Faust, 1994, Monge and Contractor, 2003]. In all analyzed networks the estimated parameter is negative indicating that the probability of observing any given tie is low. This statistic is not informative as the survey technique limits the number of citations that lead to the reduction of the graph density.

One of the mechanisms motivating social interactions is reciprocity. It appears as a significant trend for the tie emergence in coordinating and professional knowledge sharing. If two individuals were already connected by a transfer of coordinating or professional knowledge, the odds of it to become a mutual relation increase by a factor of 2.4. Positive parameter suggests that either for a more tacit as well as for a more codified knowledge knowledge a network of cooperation is characterized by the high level of cooperation and trust as a large amount of knowledge transfers happens through mutuality. The effect of mutuality is small and not significant for the object-based knowledge transfers.

Parameters estimating the power of shared partner on closure effects are negative (except for two edge-wise shared partner in the network of coordinating knowledge transfers which is positive but not significant). This indicates that all three networks of knowledge exchange do not have a tendency to cliquish structure complex extra-dyadic network dependences. The odds of triangles to close around multiple two-paths leading to a direct tie in the network reflecting coordinating knowledge transfers is very small (0.6) and not significant. Triangles do not close around knowledge hubs.

The positive Popularity spread estimate for coordinating and professional knowledge transfers indicate the tendency to centralization around few high degree nodes. The odds of having larger than randomly expected indegrees increases 8 times in coordinating knowledge and 24 times for professional knowledge transfers. The tendency towards indegree centralization is not observed in the object-based knowledge transfer network. In fact, negative estimate suggests that the odds of observing larger indegrees decreases by 0.4 times.

One of the processes influencing the emergence of knowledge transfers in the organization is a formally established order of collaborations. The email network reflects daily interactions that take place among the same set of employees. An estimator of the *email multiplexity* in a model of the knowledge transfers shows if the presence of a daily interactions influences the emergence of a knowledge sharing tie between two actors [Kleinbaum et al., 2013, Pilny and Atouba, 2018]. This term calculates the sum of independent predictor values for each tie appearing in the graph to define a structural correspondence between dependent and independent network. In the given context edge covariate term is used to access if the intensity of email communications can predict the emergence of the tie in one of the knowledge transfer networks. This configuration is positive and significant among every network indicating strong explanatory power of formal work arrangements. A strong intensity of cooperation between employees measured by email exchanges increases the odds of professional knowledge transfer by 15 times (exp(2.702)) as compared to 17 times (exp(2.829)) for coordinating knowledge transfer, and almost 20 exp(2.948) times for object-based knowledge network. Addition of multiplexity covariates significantly improves model fit.

Social exchange theory [Homans, 1958] extends the concept of tie reciprocity to account for individuals' characteristics. The probability of a tie reciprocation increases under the influence of common traits. Controlling for the endogenous effects and links alternate allows to single out an impact of these effects on a tie formation. A preferential attachment may appear as a result of actors similarities like organizational affiliation [Zappa and Robins, 2016], physical proximity [Borgatti and Cross, 2003] or gender, age, seniority [Ibarra, 1993]. Resulting clusters of similar nodes contribute to the increase in graph assortativity. Exponential random graph modeling allows fitting of several assortative mixings through estimation of attributes correlation across edges to conclude which ones rise the probability of a tie to appear.

While physical proximity is a significant configuration for all three networks, its strength varies. Being located in the same room is an important positive predictor for the tie emergence in the network of object-based knowledge transfers. The probability of observing a knowledge transfer between the colleagues of the same office increases by 12 times (exp(2.519)). Similarly strong is the effect of physical proximity on the odds of observing a tie in the coordinating knowledge transfers âĂŞ 9 times (exp(2.228)). Weaker is the effect of common location in the professional knowledge flow. The odds of the tie emergence increase by 5 times. Configuration estimating parameters for specialization homophily resulted in negative values among all three networks. Similarly in both coordinating and object-based knowledge the odds of observing the tie decreases slightly (0.5 times) between individuals belonging to the same organizational unit. Although specialization has a stronger effect in the network of the professional knowledge type, it is not significant. Together with high popularity spread it indicates that there exist a knowledge cluster concentrating requests for highly implicit information.

The effect of seniority and age difference among connected individuals is rather small. All three models reveal a slight reduction (by 1 time) in the odds of forming a tie when the distance in years served or in the age of two employees grows by 1 year. Non-significant seniority match in the object-based network suggests that similarity in organizational tenure does not affect knowledge transfers. In a similar way, the difference in age of two individuals is not a significant predictor of the professional knowledge transfer. None of the parameters for male gender homophily configuration is identified as significant. The odds of observing a knowledge transfer tie between two females in coordinating knowledge network significantly increases by 1.4 times. Although this parameter is negative for professional knowledge, its effect is small and not significant. The overall effect of gender on knowledge transfers is weak.

To validate the goodness of fit of suggested models explaining the emergence of the knowledge transfers among employees I run 100 simulations and analyze resulting networks. Suggested model shows a good fit to the observed data if observed configurations fit in the the confidence interval of simulated distribution. Visual analysis of the frequency distribution for indegree, outdegree, edgewise-shared partner statistics demonstrate that the data observed in all three networks falls within 95% CI distribution of simulated results (see Appendix C). Additionally, I control the convergence of the models through MCMC diagnostics. Appendix shows MCMC trace plotting configuration parameters for all three networks (see Appendix D). Analysis confirms that they are randomly distributed over observed values (y axis). MCMC density plots confirm that the difference between observed and simulated data for all the sample statistics is centered at zero and follows normal distribution.

5 Conclusions and Limitations

The present research called into question the significance of the knowledge nature effect on the structural difference between informal employees interactions. This paper uses exponential random graph modeling to identify crucial factors influencing knowledge transfer tie formation in a given corporate setting. Suggested models estimate the power of exogenous and endogenous factors for predicting the emergence of knowledge transfer ties. Obtained results suggest that the effects of fitted configurations differ among the three models.

Positive and significant parameters for the same physical location reveal that being located in proximity with each other increase the probability for employees to cite each other in all three analyzed networks. This effect is the strongest in the object-based knowledge network and more moderate in coordinating knowledge transfer network. Much weaker, but still significant is the impact of close location on professional knowledge sharing. This findings confirm the previous hypothesis suggested by Kleinbaum et al. [2008] about the underestimation of the predictive power of departmental boundaries for predicting knowledge exchange practices.

Specialisation homophily either reducing the probability of knowledge transfer or cease to be significant with the prevalence of the tacit component. As such, I partially confirm the first hypothesis about the crucial role of strategical workplace arrangements for operational knowledge transfers. These results are in line with previous suggestions about the crucial role of specialist group in facilitating exchange of tacit knowledge. [Davenport et al., 1998, Christensen, 2007, Blankenship and Ruona, 2009]. An example of such specialist group is a community-of-practice. Such group provides personalization strategies to facilitate knowledge sharing through a common identity, similar absorptive capacity and frequent interactions either inside or across organizational barriers [Christensen, 2007].

Positive and significant parameter for multiplexity between knowledge sharing and formal interaction networks confirms the hypothesis about a possible relationship between the electronic communication network and self-reported informal communications. For all three networks, knowledge sharing tie is more likely to appear when employees report a higher frequency of e-mail communications. These results are consistent with the previous findings and add to the literature evaluating correspondence between the networks of registered and self-reported interactions (formal and informal communicztion networks) [Hinds and Kiesler, 1995, Monge and Contractor, 2003, Rowe et al., 2007, Tashiro et al., 2010, Johnson et al., 2012].

Reciprocity has a positive effect on the probability of transferring knowledge with the highly explicit or implicit component. This effect loses significance for the knowledge constituting a more equal combination of two components (object-based). Both networks of highly explicit (coordinating) and highly implicit (professional) knowledge transfers reveal the tendency to preferential attachment mechanism [Barabási and Albert, 1999]. Few employees are visible in the network due to a high concentration of incoming ties, thus creating a vicious cycle of the stardom contributing to their role as knowledge hubs [Oldroyd and Morris, 2012]. Unlike these two networks, object-based knowledge sharing shows the tendency to decentralization - more equal distribution of the incoming ties among actors - indicating the propensity to seek contribution from various specialized knowledge [Christensen, 2007].

Significant homophily arrangements, considered together with the general tendency towards reciprocity shaping the network of prevailing codified knowledge imply that coordinating knowledge transfers are based on mutual or strong ties among individuals within specialization units. Combined with insignificant parameters of nested hierarchical triadic closures result suggests that hierarchical ordering has a higher predictive power than self-organizing mechanisms in shaping knowledge transfers networks. Significant female homophily is consistent with the prevalence of female participation in the administrative departments.

Applied multitheoretical multilevel approach [Monge and Contractor, 2003, Johnson et al., 2012, Zappa and Robins, 2016] results as a powerful methodological framework for measuring the difference between factors leading to network emergence for evaluating the effect of three types of boundaries on the knowledge transfers: specialization boundaries (business unit), physical boundaries (common office locations), and social boundaries (gender, seniority, age) [Kleinbaum et al., 2008]. Incorporating work-flow relationships as a covariate improved the fit of the model. Though the ERGM is a powerful and flexible modeling tool, it has some limitations that may hamper its utility to researchers. Among them are computational complexity, convergence and degeneracy problems appearing with the increase of the number of covariates, and inability to analyze networks over time.

Suggested modeling technique allows making a further step in the organizational analysis by accounting for the strength of ties as a predictor. But the longitudinal nature of the organizational processes reflected in relational events remains undiscovered [Shumate and Contractor, 2013, Pilny and Atouba, 2018]. Obtained models used the information about formal work arrangements extracted from the network of e-mail conversations in a limited way. I chose to reduce inherently longitudinal data to the cross-sectional image of interaction frequency. While for the matter of this research such approach allowed to test the hypothesis about the strength of ties, the future exploration should make use of the time factor and explore possible ways of identifying a taxonomy of emails.

It is worth mentioning that this research focused on the study of one particular organizational reality. As has been previously indicated by Johnson et al. [2012], Zappa and Robins [2016] such studies do not suggest nonconforming results, yet they fail to provide generalized conclusions. More empirical investigation, specifically in the knowledge-intensive and professional service settings, can help to broaden current understanding of knowledge sharing practices in the organization and contribute to drawing general conclusions to enrich evidence-based implications for organizational practitioners.

The evidence from this study supports the idea that the nature of information imposes additional complexity on knowledge transfers [Grippa et al., 2006, Johnson et al., 2012, Zappa and Robins, 2016]. Statistical inference in the knowledge sharing practices gives a better understanding of the potential areas where established organizational arrangements facilitate the exchange of knowledge. Obtained results should be used as a basis for the managerial interventions with the aim to increase employees' embeddedness in the knowledge transfers. Insights into knowledge sharing structures dictate the choice of communication strategies and help to identify preferential attachment mechanisms indicating the presence of strategically important knowledge hubs that should be encouraged to share the information with other employees.

Hypothesis tested in this research rely on the classification of operational knowledge suggested by Christensen [2007] and based on the early work of Polanyi [1962] about knowledge tacitness. As suggested by both authors, knowledge classification fits the continuum defining the prevalence of either implicit or explicit component but does not restrict the type of the knowledge to a dichotomous classification (being strictly tacit or codified). Professional knowledge relying highly on the personal expertise tend to have more tacit nature and thus require working arrangements like communities-of-practice that contribute to the development of the feeling of belonging and trust, as well as increase absorptive capacity. Being on the other extreme of the tacit-codified continuum, coordinating knowledge is more explicit. This type is the one that contributes directly to the organizational output and thus should be seen as strategically important and reflect in the well-defined set of procedures, norms, and plans.

Suggested classification allows identifying a group of knowledge transfers that require in-depth research and more sophisticated tools for modeling to reveal different mechanisms of emergence when compared to the professional or coordinating knowledge. While possessing both implicit and explicit components, object-based knowledge falls in the middle of the tacit-explicit continuum. The degree of explicitness dictates codification strategies as useful tools to share this resource. Due to the high reliance on personal expertise in identifying the "object" personalized strategies may result in more efficient in sharing.

Christensen [2007] argues that without knowing who knows what and where knowledge exists, knowledge sharing will not take place, hindering overall organizational performance. The soft nature of the knowledge sharing process stands on the way of this realization and requires evidence-based proofs to gain recognition as a factor playing an important role in developing final organizational output. At long last we know more than we can tell, we know as much as we can measure [Cross et al., 2003].

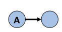
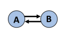
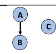
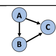
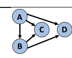
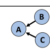
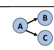
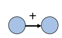
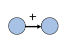
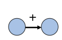
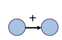
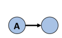
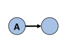
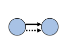
6 Appendices

*Appendix A

Survey.

1. *Which one, among your colleagues, suggested solutions linked to your working activity through already existing procedures and practices in the last 6 months?*
2. *Which one, among your colleagues, worked to translate new ideas into real applications in the last 6 months?*
3. *Which one, among your colleagues, supported you in growing the skills requested in your job during the last six months?*

Table 3: Significant Statistics for predicting knowledge transfers.

	Image	**Highly Explicit (Coordinating)**	**Explicit/Implicit (Object-based)**	**Highly Implicit (Professional)**
				Dyadic Dependent Effects - Structural Attributes
Edges		X (-)	X (-)	X (-)
Reciprocity		X (+)		X(+)
No Edge-Wise Shared Partner		X (-)	X (-)	X (-)
Friend-of-a-Friend (Transitivity)			X (-)	X (-)
Two Edge-Wise Shared Partners			X (-)	X (-)
Popularity Spread		X (+)	X (-)	X (+)
Activity Spread		X (+)	X (+)	X (-)
				Dyadic Independent Effects - Individual Attributes
Physical Proximity		X (+)	X (+)	X (+)
Department Homophily		X (-)	X (-)	
Male-to-Male				
Female-to-Female		X (+)		
Abs difference in Seniority		X (-)		X (-)
Abs difference in Age		X (-)	X (-)	
Multiplexity (e-mail network)		X (+)	X (+)	X (+)

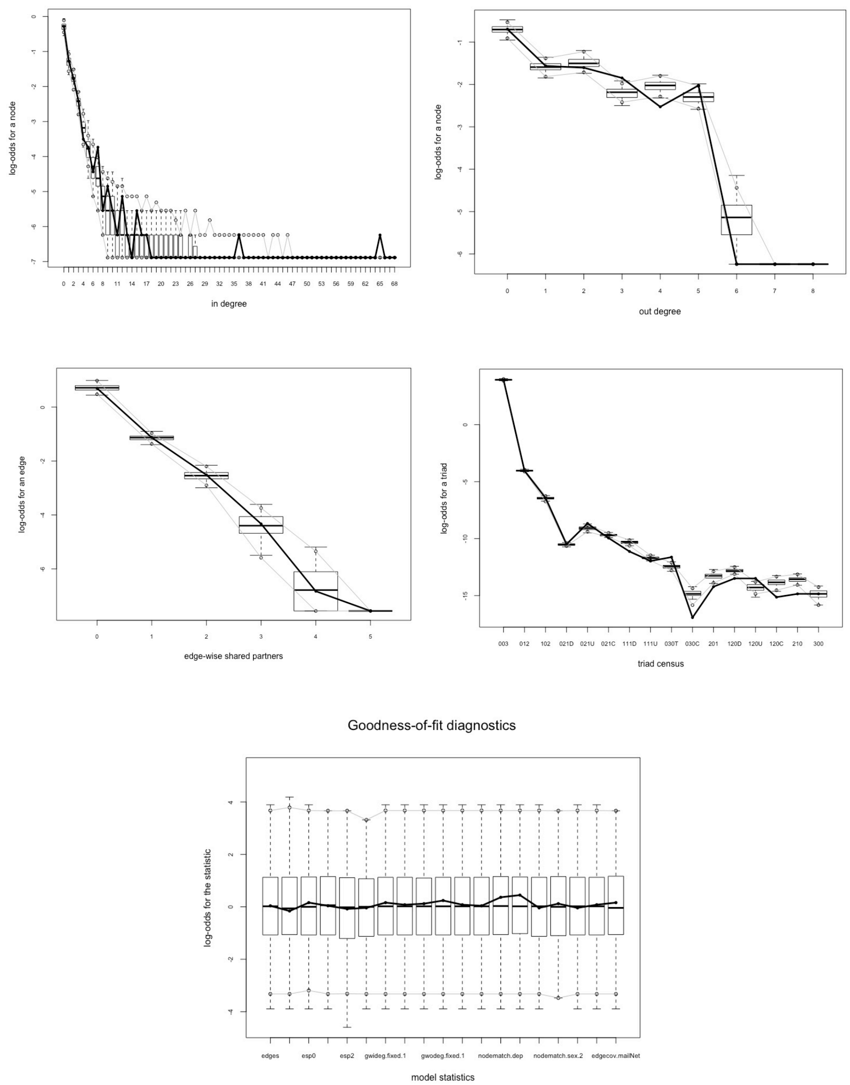
*Appendix B

Table 4: Exponential Random Graph Modeling of Knowledge Transfer Networks.

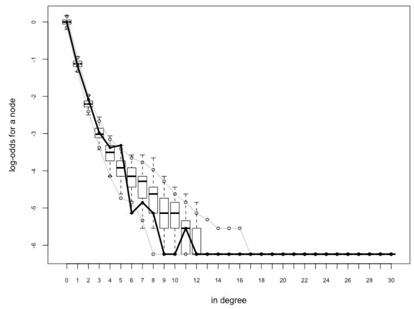
	Dependent variable: Knowledge Transfer Tie		
	Coordinating	Object-based	Professional
	(1)	(2)	(3)
Edges	-737.676^{***}	-841.630^{***}	-15.340^{***}
	(0.832)	(1.436)	(0.377)
Reciprocity	0.882^{***}	0.001	0.894^{***}
	(0.187)	(0.230)	(0.296)
No Edge-Wise Shared Partner	-0.971^{***}	-1.717^{***}	-1.675^{***}
	(0.174)	(0.152)	(0.201)
Transitivity	-0.087	-0.540^{***}	-0.380^{***}
	(0.161)	(0.116)	(0.139)
Two Edge-Wise Shared Partners	0.060	-0.467^{***}	-0.428^{**}
	(0.208)	(0.176)	(0.208)
Popularity spread ($\alpha = 1$)	2.080^{***}	-0.970^{***}	3.191^{***}
	(0.456)	(0.254)	(0.795)
Activity spread ($\alpha = 1$)	130.416^{***}	146.974^{***}	-13.694^{***}
	(0.369)	(0.476)	(0.466)
Department Homophily	-0.631^{***}	-0.677^{***}	-0.179
	(0.158)	(0.205)	(0.218)
Physical Proximity	2.228^{***}	2.519^{***}	1.635^{***}
	(0.156)	(0.200)	(0.217)
Male-to-Male	0.057	0.031	0.088
	(0.066)	(0.077)	(0.091)
Female-to-Female	0.323^{***}	0.051	-0.299
	(0.101)	(0.124)	(0.198)
Abs. difference Seniority	-0.003^{**}	-0.001	-0.004^{***}
	(0.001)	(0.001)	(0.001)
Abs. difference Age	-0.002^{***}	-0.001^{***}	-0.0004
	(0.0003)	(0.0004)	(0.0005)
Multiplexity (e-mail network)	2.829^{***}	2.948^{***}	2.702^{***}
	(0.083)	(0.097)	(0.120)
Akaike Inf. Crit.	7,890.200	5,826.763	3,555.904
Bayesian Inf. Crit.	8,089.368	6,015.448	3,744.589
Note:			*p<0.1; **p<0.05; ***p<0.01

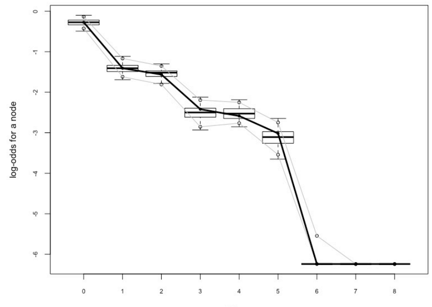
*Appendix C

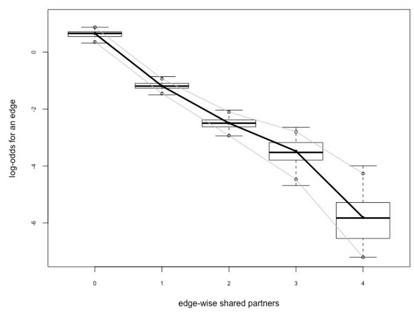
Goodness of fit diagnostic for ERGM of Coordinating Knowledge Transfers.

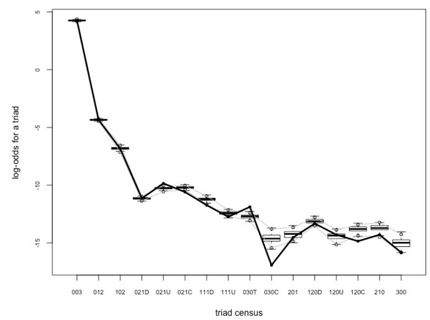


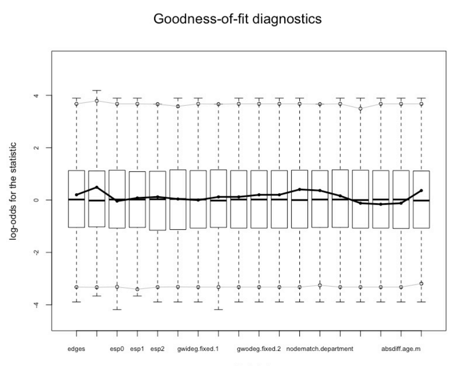
Goodness of fit diagnostic for ERGM of Object-based Knowledge Transfers.



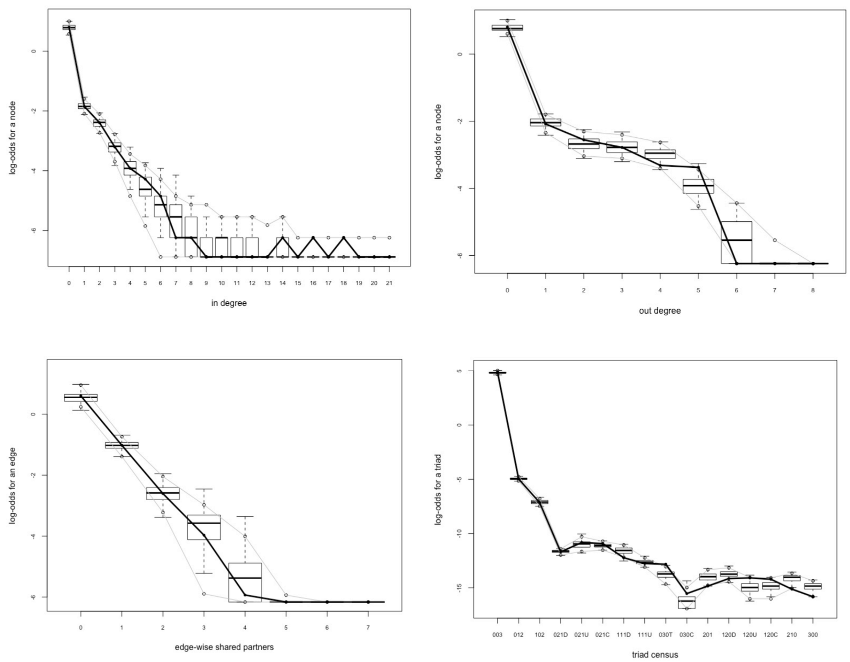


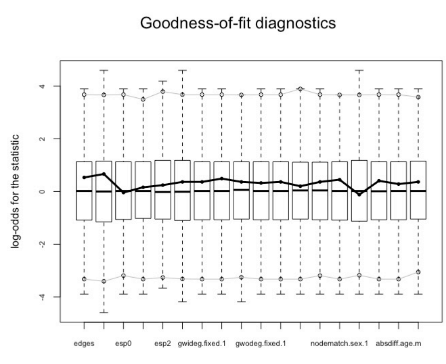






Goodness of fit diagnostic for ERGM of Professional Knowledge Transfers.





*Appendix D

Traceplots of MCMC Chain Maximum Likelihood Estimates and Distributions of Simulated Network Statistics Compared to Observed Values.

References

M. Alavi and D. E. Leidner. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1):107–136, 2001.

T. J. Allen. *Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organisation.* Cambridge, MA, 1977.

A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

S. S. Blankenship and W. E. A. Ruona. Exploring Knowledge Sharing in Social Structures: Potential Contributions to an Overall Knowledge Management Strategy. *Advances in Developing Human Resources*, 11(3):290–306, 2009.

S. P. Borgatti and R. Cross. A relational view of information seeking and learning in social networks. *Management Science*, 49(4):432–445, 2003.

R. Burt. *Structural holes.*, volume 100. Harvard University Press, Cambridge, MA, 1992.

P. H. Christensen. Knowledge sharing: moving away from the obsession with best practices. *Journal of Knowledge Management*, 11(1):36–47, 2007. ISSN 1367-3270.

L. J. Colfer and C. Y. Baldwin. The mirroring hypothesis: Theory, evidence, and exceptions. *Industrial and Corporate Change*, 25(5):709–738, 2016.

S. R. Corman and C. R. Scott. Perceived Networks, Activity Foci, and Observable Communication in Social Collectives. *Communication Theory*, 4(3):171–190, 1994.

R. Cross, S. P. Borgatti, and A. Parker. Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, 44(2):25–47, 2002.

R. Cross, A. Parker, L. Prusak, and S. P. Borgatti. *Knowing What We Know : Supporting Knowledge Creation and Sharing in Social Networks.* 2003.

R. L. Cross and A. Parker. *The hidden power of social networks: Understanding how work really gets done in organizations.* Harvard Business Press, 2004.

T. H. Davenport and L. Prusak. *Working knowledge: How organizations manage what they know.* 2000.

T. H. Davenport, D. W. D. Long, and M. C. Beers. Successful Knowledge Management Projects. *Sloan Managemenl Review*, 39(2):43–58, 1998.

M. Granovetter. The Strength of Weak Ties, 1973.

K. A. Grant. Tacit Knowledge Revisited - We Can Still Learn from Polanyi. *Journal of Knowledge Management*, 5(2):173 – 180, 2007.

R. M. Grant. Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, 17 (May):109–122, 1996.

F. Grippa, A. Zilli, R. Laubacher, and P. A. Gloor. E-mail May Not Reflect The Social Network. *SPIE Newsroom*, pages 1–6, 2006.

M. Handcock, D. Hunter, C. Butts, S. Goodreau, P. Krivitsky, and M. Morris. ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks, 2018.

P. Hinds and S. Kiesler. Communication across Boundaries: Work, Structure, and Use of Communication Technologies in a Large Organization, 1995.

G. C. Homans. Social Behavior as Exchange. *American Journal of Sociology*, 63(6):597–606, 1958.

D. R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29(2): 216–230, 2007.

D. R. Hunter, S. M. Goodreau, and M. S. Handcock. Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.

E. Ibarra. Network Centrality, Power, and Innovation Involvement : Determinants Of Technical and Administrative Roles. *The Academy of Management Journal*, 36(3):471–501, 1993.

R. Johnson, B. Kovács, and A. Vicsek. A comparison of email networks and off-line social networks: A study of a medium-sized bank. *Social Networks*, 34(4):462–469, 2012.

P. Keith. Technologies, products and organization in the innovating firm: what Adam Smith tells us and Joseph Schumpeter doesn't. *Industrial and Corporate change*, 7(3):443–452, 1998.

A. M. Kleinbaum, T. E. Stuart, and M. L. Tushman. *Communication (and Coordination?) in a Modern, Complex Organization*. Harvard Business School, Boston, MA, 2008.

A. M. Kleinbaum, T. Stuart, and M. Tushman. Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organization Science*, 24(5):1316–1336, 2013.

D. Knoke and S. Yang. *Social network analysis*. Sage, (vol. 154) edition, 2008.

G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757): 88–90, 2006.

D. Krachardt and J. Hanson. Informal networks: the company behind the chart. *Harvard Business Review*, 71(4):104–111, 1993.

D. Krackhardt. Constraints on the interactive organization as an ideal type. In C. Heckscher and A. Donnellon, editors, *The Post-Bureaucratic Organization: New Perspective on Organizational Change*, pages 211–222. Sage Publications, Thousand Oaks, CA, 1994.

R. K. Moenaert and W. E. Souder. An analysis of the use of extrafunctional information by R&D and marketing personnel: review and model. *Journal of Product Innovation Management*, 7(3): 213–229, 1990.

P. R. Monge and N. S. Contractor. *Theories of communication networks.* Oxford University Press, USA., 2003.

I. Nonaka. The knowledge-creating company. *Harvard business review*, 1991.

I. Nonaka and T. Hirotaka. *The knowledge creating company: how Japanese companies create the dynamics of innovation.* New York, 1995.

I. Nonaka and N. Konno. The concept of "Ba": Building a foundation for knowledge creation. *California management review*, 40(3):40–54, 1998.

J. A. Odero and W. Otuya. Critical Review of Literature on Knowledge Management Strategy and Organizational Performance. *International Journal of Management and Commerce Innovations*, 5(2):741–748, 2017.

J. B. Oldroyd and S. S. Morris. Catching falling stars: A human resource response to social capital's detrimental effect of information overload on star employees. *Academy of Management Review*, 37(3):396–418, 2012.

F. O. Omotayo. Knowledge Management as an important tool in Organisational Management: A Review of Literature. *Library Philosophy and Practice*, 2015.

P. E. Pattison and G. L. Robins. *Probabilistic network theory. Handbook of Probability Theory with Applications.* Sage Publications, Thousand Oaks, CA, 2008.

A. Pilny and Y. Atouba. Modeling Valued Organizational Communication Networks Using Exponential Random Graph Models. *Management Communication Quarterly*, 32(2):250–264, 2018.

M. Polanyi. *Personal Knowledge: Toward a Post- Critical Philosophy.* Harper Torchbooks, New York, 1962.

M. Polanyi. *The tacit dimension.* Garden City, NY, 1967.

E. Quintane and A. M. Kleinbaum. Matter Over Mind? E-mail Data and the Measurement of Social Networks. *Connections*, 31(1):22–46, 2011.

G. Robins. A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, 57(6):261–274, 2013.

R. Rowe, S. J. Stolfo, and R. Rowe. Segmentation and Automated Social Hierarchy Detection through Email Network Analysis Automated Social Hierarchy Detection through Email Network Analysis. (January), 2007.

M. Shumate and N. Contractor. Emergence of multidimensional social networks. In L. Putnam & D. K. Mumby (Eds.). In *Handbook of organizational communication*, pages 449–475. Thousand Oaks, CA: SAGE, 2013.

T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36(1):99–153, 2006.

G. Soda and A. Zaheer. A network perspective on organizational architecture: Performance effects of the interplay of formal and informal organization. *Strategic Management Journal*, 33(6):751–771, 2012.

H. Tashiro, J. Mori, N. Fujii, and K. Matsushima. Email Network Analysis for Organizational Management. (June 2009):958–963, 2010.

W. Tasi and W. Tsai. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit, innovation and performance. *Academy of Management Journal*, 44(5):996–1004, 2001.

J. D. Thompson. *Organizations in Action: Social Science Bases of Administrative Theory.* Transaction Publishers, New Brunswick, NJ, 1967.

S. Wasserman and K. Faust. *Social network analysis: Methods and applications.* Cambridge university press, (vol. 8) edition, 1994.

P. Zappa and G. Robins. Organizational learning across multi-level networks. *Social Networks*, 44: 295–306, 2016.

Modelling longitudinal effects of communication dynamics on the emergence of informal leaders.

Anna Velyka [1,3]

[1] Despina, Big Data Lab

[2] Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Italy

[3] Aizoon

Abstract

This paper examines the structural similarity between the network of registered and self-reported interactions. Organizational studies traditionally rely on self-reporting methodology to identify informal leaders. This approach is subjected to response bias and limits the dynamical nature inherent to communications structure to a static representation. The literature investigating alternative sources of data for the analysis of informal leadership positions is scarce. This paper addresses the gap through a comparative study of the intraorganizational interactions originating from two different sources: self-reported questionnaire and electronic communications. History of e-mail communications provides information about individual trajectories of growth within the organizational network of collaborations. Through the use of structural equation modeling, I determine the effect of the time-related change in weighted degree, betweenness and closeness centrality as factors related to the emergence of informal leadership position.

In line with the previous research obtained results conclude that the structure of interactions observed in the e-mail network does not provide an exhaustive approximation of informal interactions between employees. I find that the growth of betweenness and closeness centrality in the network of electronic communications does not show significant significant relation with high centrality in the network of informal interactions. However, the results show that the growth in weighted degree centrality exercises a significant effect over the definition of informal leadership. The results provide important implications for the analysis of employees performance in the organizational setting suggesting a methodological limitations to the study of employees' involvement in the organizational structure and alternative to carrying out a survey.

Keywords: informal leadership; longitudinal effects of communications; latent growth modeling;

1 Introduction

Technological advances brought by incorporation of electronic communication tools into the operational workflow is a leading cause of structural changes in organizational mechanisms. Constraints of time and distance are losing significance [Fulk and DeSanctis, 1995], contributing to the elimination of barriers between organizational actors. Better connected employees enjoy higher returns from interactions [Burt, 2002], implicitly supporting the uniqueness of competitive advantage and efficient organizational functioning.

The data extracted from electronic communication tools is an indicator of employees connectedness in the information exchange networks [Ducheneaut and Bellotti, 2002, Kleinbaum et al., 2008, Quintane and Kleinbaum, 2011]. Automatic processing of e-mail interactions reveals analytical power preserved in the longitudinal nature of communication data. Statistical analysis of social processes from e-mail interactions uncovers the dynamics of social linkage between organizational actors [Tashiro et al., 2010], displays patterns of individual connectedness, mutual acquaintances and recognition [Burt, 1992].

Previous research relying on the analysis of the e-mail communications focuses on the definition of the mismatch between the network constructed from the e-mail data and other registered or self-reported data [Monge and Contractor, 2003, Kossinets and Watts, 2006, Kleinbaum et al., 2008, Quintane and Kleinbaum, 2011, Johnson et al., 2012]. Understanding the structural difference between two sources of data is essential for determining the value of electronic interactions within the framework of organizational research. Actual research extends beyond the study of conformity of the static and dynamic networks to verify the significance of centrality correlations. This study investigates the effect of growth in centrality trajectories observed in the network of electronic communications on the emergence of informal leadership positions. The analysis relies on data about employees performance in knowledge sharing networks, gathered through the survey and the logs of email communications for two years before the beginning of the study. This paper applies Structural Equation Modeling to estimate the direct effect of latent variables summarizing communication dynamics through intercept and slope of individual change on the observed variable of interest representing informal leadership.

This paper aims to contribute to the growing body of evidence-based research on electronic communication claiming its defining role in the analysis of social relations inside the organization. Structural analysis of technology-organization relationship offers a theoretical base for the broader understanding of individual performance [Fulk and DeSanctis, 1995, Kleinbaum et al., 2008]. Obtained results support the previous claim that the data collected by parsing electronic mail and surveying employees should be used to answer different questions of organizational research [Quintane and Kleinbaum, 2011]. Leadership positions identified through e-mail data partially resemble the self-reported social structures. As survey methodology can be labor- and time-consuming, e-mail interactions must be further researched as a proxy of collaboration and organizational knowledge exchange. This analysis suggests implications for extending a methodological approach for the evaluation of employees performance and developing more flexible tools for identification of leadership positions.

2 Theoretical Background

Organization functions through a set of procedures, rules, and programs. Such arrangements are well-defined in a formal structure that dictates delegation of authority and relationship among various hierarchical levels. Postindustrial organizations strive to develop participative culture [Fulk and DeSanctis, 1995] where individuals are motivated to cooperate. Interpersonal relations mandated by hierarchical order manifest itself in the channels officially established for organizational communications. E-mail interactions offer an approximation of frequencies of daily collaborations [Ducheneaut and Bellotti, 2002, Kleinbaum et al., 2008, Quintane and Kleinbaum, 2011]. Several studies suggested that the structure of collaboration patterns within e-mail network provides a basic understanding of informal communities [Tyler et al., 2003, Diesner and Carley, 2005]. However, the literature does not offer empirical analysis for the identification of informal leadership positions in the network of e-mail collaborations.

Daily cooperation does not always follow formally established links. Organizational members that develop informal relationships based on preferential attachment chose communication channels independently. Interdependencies emerging on the grounds of such intrinsic motivation provide insights into advice seeking, expertise sharing, and knowledge transfer [Krachardt and Hanson, 1993, Zappa and Robins, 2016]. Referred to as "neutral social interactions" in Allen [1977] these interactions indirectly contribute to organizational output and signaling individuals a potential trove of information resources [Kleinbaum et al., 2013]. The informal network offers an overview of the actual knowledge exchange processes underlying operational activities [Krachardt and Hanson, 1993, Monge and Contractor, 2003, Agneessens and Wittek, 2012].

As informal networks evolve, the most actively participating employees acquire a new role of informal leadership. This structure does not exist in isolation, but rather as complementary to individual behavioral practices resulting from hierarchical norms. Implicit nature of informal communications creates the difficulty for identifying the data suitable for the analysis. With the development of the social network, analysis practitioners obtained a methodological tool that allowed to recreate the structure of the informal relationship [Wasserman and Faust, 1994, de Toni and Nonino, 2010]

Qualitative methods as questionnaires, interviews, and surveys increasingly being applied for gathering the data on informal social networks embedded in organizational context [Marsden, 1990, Odero and Otuya, 2017]. This methodology requires direct interaction with the subject under study as interactions are self-reported. A researcher formulates the question based on the nature of the domain of interest. Individuals present their vision of communications in this domain by selecting the members of the sample from the name generator list. Active approach for social data gathering provides flexibility to the researcher in defining the content of the question, duration of the reference period, geographical and hierarchical boundaries.

One of the most significant distinctions of a survey from other methods of data collection is the possibility to estimate the content of interaction [Quintane and Kleinbaum, 2011]. The main critique associated with these methods refers to reliance on individual judgments, which exposes results to subjectivity and bias [Kannan and Aulbur, 2004]. Reliance on individual capacity to recall communication activity can result in incomplete data and thus distorted reflection of the communication network. Individuals are more likely to cite peers with whom they have a more stable and frequent relationship [Freeman, 1978], or peers that are more visible in the organizational network [Brewer, 2000].

Analysis of social interaction based on the data obtained through parsing the logs of electronic communication relies on a passive approach to data collection as it does not build on respondents participation. The nature of technology allows extracting communication records from servers ensuring reliable and complete information [Fulk and DeSanctis, 1995]. The growing popularity of organizational network analysis of electronic communication is attributed to the intrinsic features of the technology that lead to changes in the organizational forms. Among these features is the rise in the quantity of information that is simultaneously traveling along corporate channels contributing to the increase of the speed of interactions and increasing connectivity [Fulk and DeSanctis, 1995].

In the majority of organizational contexts, email has become a medium for formal communications and the principal channel for the workflow system. Recent research concluded that at least in some organizations, patterns of e-mail interactions are similar to those of face-to-face, telephone, meetings, and calendar data [Grippa et al., 2006, Kleinbaum et al., 2008]. These assumptions suggest that social interactions observed in the email communication are an approximation of the workflow system in the organizational setting. Organizational network analysis based on the data from e-mail communications is unobtrusive and involves lower costs of data collection [Quintane and Kleinbaum, 2011]. However, this analysis poses a methodological difficulty. Fewer theories and methodological frameworks exist for the study of longitudinal data on communications, and thus it requires more sophisticated analytical skills [Ancona et al., 2001, Quintane and Kleinbaum, 2011].

An issue of correspondence between self-reported and observed interactions has been previously observed in the literature [Brewer, 2000, Marsden, 2005, Quintane and Kleinbaum, 2011]. Early attempts to define the correspondence between communication structures constructed from the recall data (survey) and direct observation (registered communications) established the absence of "acceptable accuracy" and concluded that "[...]recall of communication links in a network is not a proxy for communication behavior[...]"[Bernard et al., 1981, 11]. One of the inherent characteristics defining the difference between the two structures is the source of data. The logs of email exchanges expect to observe a dynamic social network evolving over time. The data extracted by observing or surveying individuals offers a one-time measurement, static perspective on the network.

Another critical difference manifests itself at the stage of data extraction. Email data is preserved in a registered manner, thus providing insights into actual interactions taking place between the subjects. Survey-based data represents a reflection, conditioned to the observer who is registering information and respondent who is providing the view over the communication activity. The difference between active and passive reflection on interpersonal communications goes beyond the data extraction methodology.

Survey data offers a specific perspective on communication. This is a flexible methodology, giving a possibility to decide which type of relationship to focus on (including the nature of the relationship and also the type of the channel used). Email data summarizes social connections with a limited perspective over the nature of interactions taking place within one channel.

The distinction manifests itself even in the underlying structural attributes of two networks. Hidden transitive nature in the survey measure is explained by popularity-based centralization, and various actors attribute as hierarchy, tenure, gender. As respondents present their estimation of a communication network, it reflects independent relations. The clustering in the email network is based on structural homophily of different actor attributes (especially departmental and spatial proximity. Email data usually does not imply independence since the reciprocity of ties is rather high. Reciprocity of survey and email is structurally different.

Understanding, the connection between these two sources of communication data, is essential for determining the value of electronic data for organizational research. This paper makes an inquiry into the nature of the connection between electronic communication activity of individuals and their rise to recognition as the key expertise holders. The analysis shifts the discourse from the correspondence to causation to understand systematic differences between the email and the survey-based networks.

3 Methodology

Analysis of continuous relational data requires complex modeling tools [Quintane and Kleinbaum, 2011]. Existing research analyzing electronic communications is limited and does not agree on a common theoretical and methodological approach. Preprocessing of longitudinal data relies on the aggregation of time window and definition of the link. Suggested techniques for the identification of the appropriate time window vary among the nature of data, research question, and methodological approach. The window size defined for the analysis of the telecommunication data used in human social network analysis varies from one or several days [Aiello et al., 2000, Farine, 2018] to weeks [Nanavati et al., 2006, Sims et al., 2014], from one month [Seshadri et al., 2008, Tashiro et al., 2010, Melorose et al., 2015] to several months [Kossinets and Watts, 2006, Grippa et al., 2006, Onnela et al., 2007, Lambiotte et al., 2018, Quintane and Kleinbaum, 2011, Krings et al., 2012]. Researchers conclude that the sample size and the definition of the aggregated time window play a crucial role in the description of underlying factors defining communication structure evolution.

The nature of electronic channel for communication allows reaching out simultaneously to an unlimited number of actors. The data of e-mail logs is noisy as it contains mass mailings. To reduce the noise present in the data set analysis of electronic communications requires an approximation of the link. A common approach to the definition of socially meaningful relationship is to reach lower network density by reducing the number of messages with multiple recipients. The threshold for identification of mass mailings varies among researchers, guided principally by practical consideration and can range from 2 recipients [Johnson et al., 2012], 4 recipients [Kossinets and Watts, 2006, Kleinbaum et al., 2008, Quintane and Kleinbaum, 2011], to 5 recipients [Melorose et al., 2015].

A common approach applied in the organizational studies for the analysis of time-stamped communication logs is to aggregate it into static network [Quintane and Kleinbaum, 2011], thus expanding the focus of the investigation to a new variable - weight, which represents the frequency of interactions (links) between actors (nodes) within defined time window [Krings et al., 2012]. A static snapshot of longitudinal network disregards its fundamental property - time-related change. Observing changing patterns of individual interaction within the overall social structure can give a more accurate estimate of behavioral change leading to the emergence of reputational characteristic [Ibarra, 1993].

Longitudinal perspective captures the change in interaction structure through multiple measurements performed over selected time intervals. The literature offers several methodological frameworks to estimate the effect of the change in data with several waves of measurement [Breitsohl, 2018]. Analysis of variances (ANOVA) provides a reliable framework for statistical interpretation of change through a comparison of group means over time [Breitsohl, 2018]. While being the most widely used design, ANOVA is sensitive to the presence of missing values and does not allow fitting time-varying covariates and non-orthogonal predictors. Unlike ANOVA, the multilevel approach allows modeling correlations between repeated measures. Multilevel models for repeated measures offer greater statistical power for fitting both fixed and random effects. The design is further expanded to fit time as the independent predictor at the lowest level of analysis with no restriction on the number of time-varying covariates [Hox and Stoel, 2014]. However, the study of change as the predictor variable within the multilevel framework limits estimation to only directly measured predictors and offers biased results as the predictors are measured without errors [Hodis and Hodis, 2013].

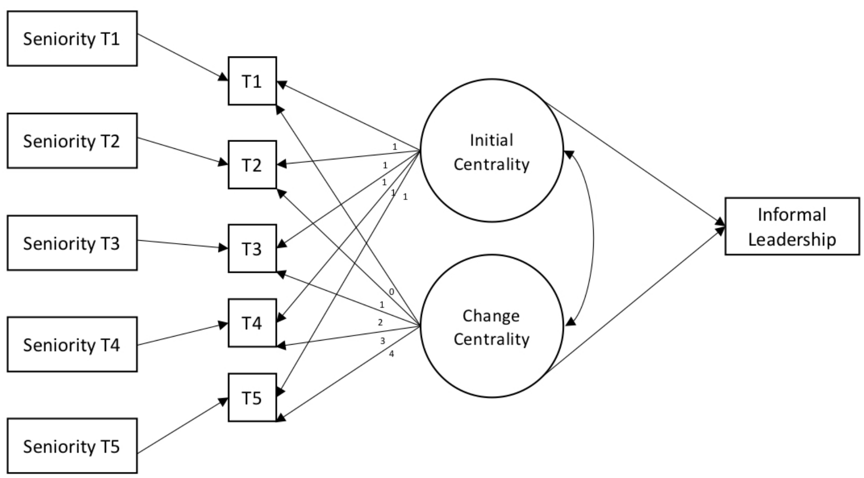
An alternative approach to performing the longitudinal analysis is Structural Equation Modelling (SEM). This approach offers a more flexible statistical design for the study of time-varying predictors. Latent growth modeling (LGM) within the SEM framework permits the use of multiple indicators of a latent construct and explicitly models measurement error in predictors as well as the outcomes. Compared to multilevel modeling approach, LGM does not impose restrictions on the presence of missing data and offers a diverse selection of goodness of fit statistics [Hodis and Hodis, 2013]. As a tradeoff for increased flexibility SEM imposes assumptions on the size of the sample (>200), the least required number of repeated measurements (>3) and expects equal time spacing across measured objects [Breitsohl, 2018]. The analysis fits only continuous measures of dependent variables and expects a linear change in the data.

3.1 Model

To examine change in the communication activity of the organizational actors as a latent factor this research relies on Structural Equation Modeling. The design of the study consists of two steps. Latent growth model fits the data on centrality statistics calculated from the e-mail network (degree, betweenness, closeness) at each selected time window to define individual curves of change. The output of the latent growth model represents the single trajectories of communication activity estimated through the values of corresponding intercepts and slopes. The model further tests the effect of two latent variables on the dependent variable defining informal leadership.

Figure 1 illustrates the path diagram for the estimation of the relationship between directly observed variable defining informal leadership and unobserved variables measuring the evolution of individuals communication activity in the e-mail network. Unobserved or latent variables are represented as circles and defined by factor loadings corresponding to 5 measurement time windows summarising the intercept and the slope of individual change trajectory. The loadings of intercepts are set to 1 to model the start point for every employee. The loading of the first time point of the slope is set to 0 with the next number approximating the time difference successively from 0 to 4 corresponding to every six months. The length of organizational tenure (graphically illustrated as a rectangular) is a control factor for the growth in the actors' centrality. The curved arrow between latent variables indicates covariance between the intercept and slope factors. Two lines connecting a latent intercept and slope to the observed variable defining informal leadership indicate the direction of the expected causal effect.

Figure 1: Graphical representation of the model.



Individual position in the network of workflow communications define organizational status, involvement in information flows, and control over corporate resources [Ibarra, 1993]. I extend the previous works of Ibarra [1993] and de Toni and Nonino [2010] and use three network centrality characteristics calculated from the graph of e-mail conversations for this analysis: weighted degree, betweenness, and closeness centrality. *Degree centrality* identifies the strength of the node and the importance of an employee's formal leadership position in the organization. Calculated as a sum of the number of links that connect the node to others this statistics shows central figures who can potentially exercise greater control over the accumulation of information resource and influence others [Yan and Ding, 2009].

Betweenness centrality approximates brokerage role of the node to determine which employees eventually grow to support information exchange among groups. This statistic measures the extent to which a node lies between other nodes on their shortest paths. Highly central employees facilitate the transfer of the corporate resources among others and as such obtain higher influence over the flow of informationKaye et al. [2014]. *Closeness centrality* identifies nodes that can quickly reach out to others [Kaye et al., 2014] representing more independent employees with specific knowledge resources. This centrality calculates the shortest distances from the node to all others [Yan and Ding, 2009] and depicts employees that can quickly reach out to colleagues. Higher ranks of the closeness centrality imply stronger involvement in the organizational network.

These centrality measures provide an overview of employees behavior within the organizational role. Reputational measure expressed as an informal leadership position is calculated from the network of organizational information transfers.

3.2 Data

This paper analyzes data collected in a medium size company working in the sphere of information technology consulting. The dataset originates from two sources: sociometric questionnaire and logs of e-mail messages among the employees. Two adjacency matrixes summarise interaction dynamics extracted from questionnaire responses and e-mail communications, with rows and columns corresponding to the names of employees. The values (0/1) in the cells of the matrix used to construct survey graph demonstrate the presence of informal interaction. The values in the matrix representing the e-mail graph indicate the strength of communication - the number of exchanged messages. The analysis of correspondence between the networks relies on the centrality statistics calculated from two constructed graphs.

The purpose of the questionnaire was to gather the data about the informal interactions between employees in the working communications domain. The study had an elevated response rate (93%) and resulted in 491 responses. Each employee answered three questions in order to identify three networks comprising an information network of the working domain. Figure 2 illustrates the composition of an information network of working communications within the organization previously defined by Krachardt and Hanson [1993], de Toni and Nonino [2010]. Respoondents were asked the following questions:

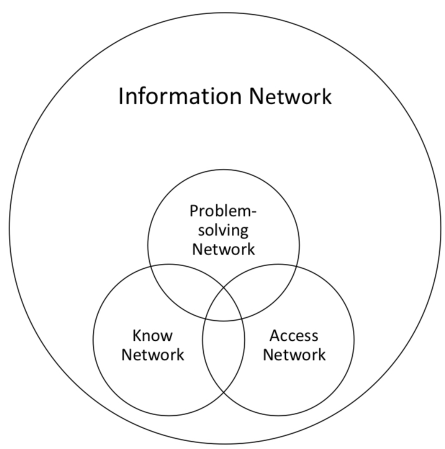
Which one, among your colleagues, suggested solutions linked to your working activity through already existing procedures and practices in the last 6 months? - Access Network.

Which one, among your colleagues, supported you in growing the skills requested in your job during the last six months? - Know Network.

Which one, among your colleagues, worked to translate new ideas into real applications in the last 6 months? - Problem-solving Network.

Problem solving network describes the advice sharing interaction to retrieve information concerning new issues in the working activity. **Access network** determines accessibility to knowledge in the organization through interactions that assist in solving an existing problem. Interactions in the **Know network** summarise the spread of experience and expertise within the organization and defines central figures who are the most qualified to answer questions about the working issues. The graph constructed from the survey united all three networks and accounted for 514 nodes (including people who were citated but didn't fil in the survey) with 1410 edges.

Figure 2: Working domain information networks



The logs of electronic communications are used to envision formal workflow in the organization. The data is available for the period from January 1, 2015 until June 30, 2017 (before the survey introduction). Interactions are aggregated in five observation windows summarising six months to examine a change in communication patterns. The analysis is limited to organizational boundaries.

All the edges representing communications directed toward actors outside the organization are deleted. In line with the previous research mass mailings threshold is defined at the level of 4 recipients [Kossinets and Watts, 2006, Kleinbaum et al., 2008, Quintane and Kleinbaum, 2011].

The total number of organizational actors at the beginning of the survey was 514 people. The population selected for this study was limited to the employees from operational departments and contained 300 employees. Administrative and managerial staff was excluded from the study to eliminate outliers in centrality statistics. Women comprised 27% of the sample. Sample age composition ranged from 25 to 67 years with the average age of 41 years. The average organizational tenure at the beginning of the survey was 6 years. 20 employees from the sample had the longest seniority of 10 years.

4 Analysis

Network defining informal leadership is built from the binary adjacency matrix reflecting self-reported impressions of the actors in the study. This network is static. Network reflecting daily e-mail interactions between subjects offers a dynamic overview of operational communications. Social network analysis results in three centrality statistics calculated as the number of links directed to and from a given node, the number of the links that are passing through it, and the average distance to all other nodes. As previously defined, obtained estimations reflect actors status and involvement and are used as predictors of informal leadership in the final model.

The study of correspondence between the network of self-reporting and electronically registered interactions starts with the correlation analysis between centrality statistics presented in Table 1. Weighted degree centrality calculated in the e-mail network is significantly and positively correlated with degree centrality in the survey network. The results are robust to the selected threshold preserving significance in the network with and without mass mailings. In particular, the e-mail network shows a more significant correlation with Problem-solving and Access networks. Due to the fewer number of citations, Know network exercises lower significance power over the probability to result central in the overall information network (0.59). For the same reason correspondence between centrality in the e-mail network and Know network is the lowest (0.34 for mass mailings and 0.31 for the reduced e-mail network).

Correlation results for betweenness centralities are inconsistent. In general correlation coefficients for betweenness centralities in the survey and e-mail networks are low. Bridging roles in the e-mail network containing mass mailings does not reveal significant correspondence with the same roles in the Problem-solving network. With the decrease in the density of electronic communication, Problem-solving network gains (0.33) and Know Network losses significance (0.069). Correlation coefficients for betweenness estimation among survey networks are positive and highly significant. Bridging roles in the Access network are more likely to correspond to bridging roles in the Know network (0.63).

Reducing the density of the e-mail network has a significant influence over closeness centrality. Closeness centrality in the aggregated Information network does not correspond to closeness centralities in the mass e-mail network (0.04) and gains significance with the decrease of density (0.22). Correlation coefficients defining closeness centrality correspondence in the e-mail network and information network are positive and significant, yet the numbers are low.

Table 1: Correlations of Centrality Statistics.

Weighted Degree Centrality	2	3	4	5	6
1 Email Network	0.99***	0.69***	0.67***	0.56***	0.34***
2 Email Network (recipients<4)		0.65***	0.63***	0.52***	0.31***
3 Information Network			0.94***	0.79***	0.59***
4 Access Network				0.61***	0.37***
5 Know Network					
6 Problem-Solving Network					

Betweenness Centrality	2	3	4	5	6
1 Email Network	0.39***	0.12**	0.071	0.20***	0.11*
2 Email Network (recipients<4)		0.25***	0.33***	0.15***	0.069
3 Information Network			0.55***	0.53***	0.69***
4 Access Network				0.31***	0.39***
5 Know Network					0.63***
6 Problem-Solving Network					

Closeness Centrality	2	3	4	5	6
1 Email Network	0.34***	0.04	0.056	0.20***	0.17***
2 Email Network (recipients<4)		0.22***	0.24***	0.26***	0.23***
3 Information Network			0.87***	0.64***	0.40***
4 Access Network				0.59***	0.41***
5 Know Network					0.49***
6 Problem-Solving Network					

Significance codes: *** 0.001, ** 0.01, * 0.05

I begin the analysis by estimating the significance of the effect of change in the centrality statistics calculated from the electronic communication graph. Suggest path for selection of appropriate latent growth model (see Appendix 1) starts with the most restrictive design constraining intercept variances to zero (Model 1). This model estimates the mean difference between individuals not allowing individual trajectories to differ and forcing residuals to be equal along five analyzed waves. For all three centralities examined this model gives a poor fit as the sample contains employees with different duration of organizational tenure, assuming that they have different estimates of centrality metrics at the beginning of the study. Model 2 allows individual statistics to start at various points and as such intercept variance is estimated. The variances are low among all centralities indicating that individuals on average do not differ significantly at the beginning of the study. The significance of the comparative fit index (CFI) for all models proves the importance of allowing actors to vary along the starting point when explaining the change in the centrality statistic.

Model 3 test the random effect of the slope. By adding a slope parameter but forcing it to be zero the model excludes slope average and estimates only the variance across actors. Covariance of the slope with the intercept is set to zero to imply that these two unobserved variables are unrelated. Suggested design still keeps residuals at the same level. Model fit does not improve for the closeness centrality and shows a very small improvement (weighted degree centrality CFI = 0.02, betweenness centrality CFI = 0.01). These results imply that the slope is consistent across selected time intervals and there are no outliers in the individual trajectories.

Restrictions for the slope and covariance introduced in the random slope model are deleted in the Model 4 while residual variance remains constrained. This design gives the actual estimation of the slope. The variance of the slope for every measurement point is significant for all three centralities. While the estimation of the slope is positive, fitted covariance results negative. These results are consistent with the assumption about dynamics in the interaction centralities. Positive estimate of the slope suggests that with the time the centrality of the actor is rising. But since the covariance is negative actors who start with higher centralities will grow slower as the ones who enter with lower starting points.

The final latent growth design (Model5) fits totally unconstrained model estimating random intercept and slope among individuals and allowing residuals to vary. The variance of residuals for all centrality statistics is decreasing from Time 1 to Time 5. A good fit strives for relatively small and equal residuals for all five waves of measurement. The estimates of residuals in this design reveal that along the five ways corresponding to two years of electronic message exchange the variance between employees' centrality statistics is decreasing. This result suggest that selected models do not violate the sphericity assumption (as the slopes rise, the estimates are coming together).

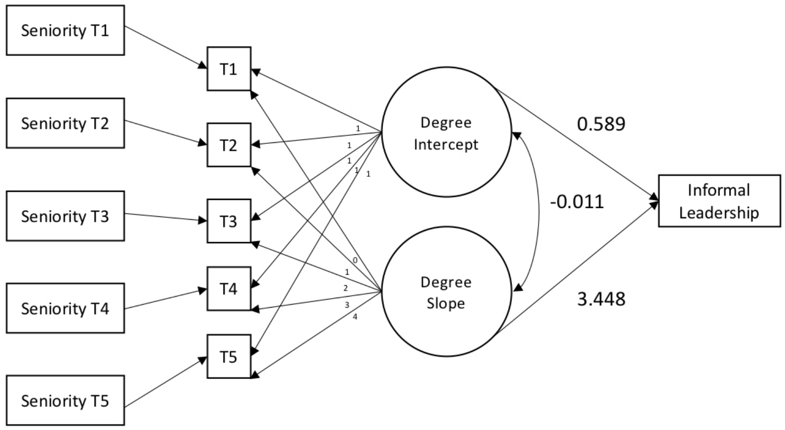
Suggested path for model selection shows an effective way of fitting change in centrality statistic as the residuals in the model decreases with the adding of new parameters. The change in the interaction activity of employees requires an estimation of the random slope and intercept. The time-varying difference in the weighted degree and betweenness centrality is better explained with the unconstrained residuals model (Model 5) as the variance among individuals grows at the time 5 and time 4 (correspondingly). Model 4 gives a better fit for the explanation of the closeness centrality as the variances among individuals are consistent over time. Thus, the change in this centrality can be explained either with the average or individual variances.

The main assumption made in this paper is that the individual trajectories of change observed in the network of work-flow interactions significantly influences the emergence of informal leaders in the organizational structure. Conducted latent growth modeling suggests that linear change in centrality statistics is explained by fitting two unobserved variables - intercept and slope. At the final stage of analysis, I compose three structural equation models estimating the effect of change in weighted degree, betweenness and closeness centrality on the centrality statistic in the information network.

I add seniority variable to control for the rising variance in the residuals that appeared in the latent growth analysis of weighted degree and betweenness centrality. Adding this control to the model estimating the effect of closeness centrality also resulted in a decrease of residuals. Overall all models show a good fit, suggesting the adequacy of the modeled data for the analysis of the causal effect between centralities of the observed networks.

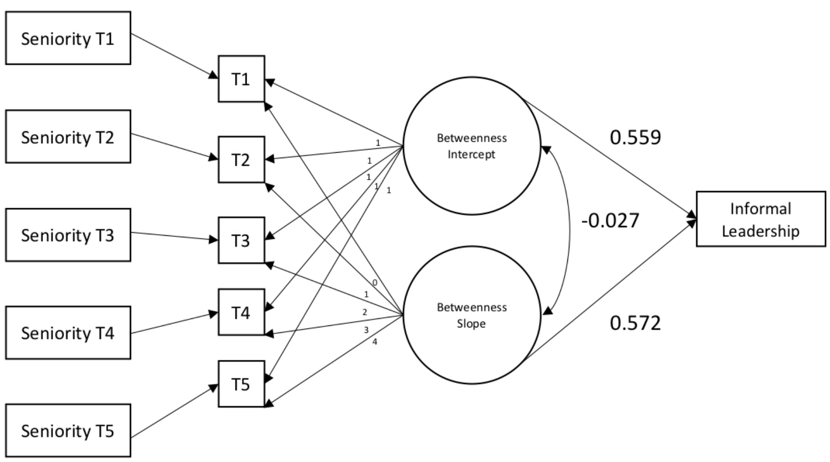
The importance of employee reflected through the weighted degree centrality in the e-mail network results as the significant predictor of the emergence of informal leadership position. Figure 3 illustrates that the estimates of the intercept (0.589) and the slope (3.448) are significant. Consistent with initially fitted latent growth model positive estimate of the slope variance and negative covariance imply that the variances in the individual trajectories tend to decrease. In line with the results of the correlation test, R squared gives a high estimate, suggesting that latent variables defining the strength of the node over time explain 36% of the variation in the informal centrality measure.

Path diagram in the figure 4 reveals the causal influence of the betweenness centrality in the electronic communication network over the informal centrality. While the intercept of the centrality statistics modeled as a random variable and reflecting single starting point results as a significant predictor, the slope reflecting the change over time is not significant p=0.512. The R^2 is rather small, accounting only 24 % of the variation in the informal centrality.



Goodness of fit	DF(X^2) $p < 0.001$	RMSEA < 0.08	SRMR <0.08	CFI ≥0.90	R^2
Results	35(226.522)	0.138	0.044	0.938	0.357

Figure 3: Path diagram for the causal effect of Weighted Degree Centrality on Informal Leadership.

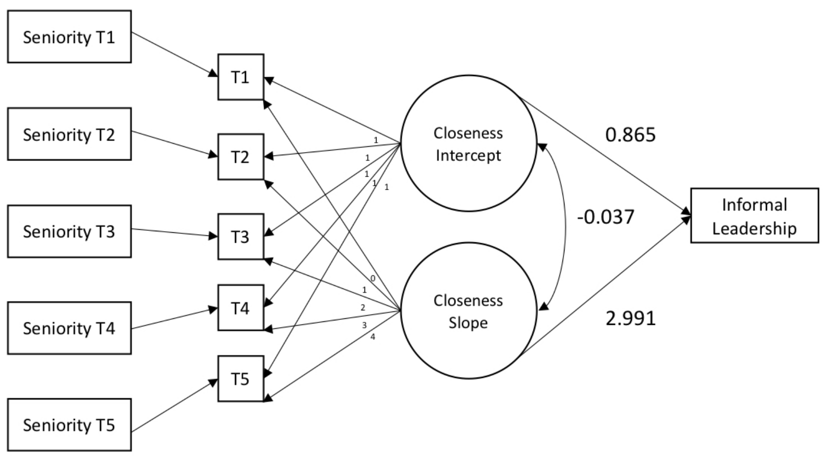


Goodness of fit	DF(X^2) $p < 0.001$	RMSEA < 0.08	SRMR <0.08	CFI ≥0.90	R^2
Results	35(79.511)	0.067	0.040	0.968	0.246

Figure 4: Path diagram for the causal effect of Betweenness Centrality on Informal Leadership.

Higher starting positions and the change in the actors involvement reflected through closeness centrality exercise significant effect over the informal leadership (See Figure figure 5). The estimate of the intercept is slightly higher than in the weighted degree (0.865) and the estimate of the slope is 2.991, both highly significant. This model shows the best goodness of fit statistics, yet results in a small R^2 value (0.243), accounting for the same number of variance as the brokering centrality.

Figure 5: Path diagram for the causal effect of Closeness Centrality on Informal Leadership.



Goodness of fit	DF(X^2) $p < 0.001$	RMSEA < 0.08	SRMR <0.08	CFI ≥0.90	R^2
Results	39(38.845)	0.000	0.034	1.000	0.243

5 Conclusions

The present study was designed to further investigate the correspondence between electronically registered and self-reported organizational interactions. In this investigation, I assessed the causal relationship between centrality statistics calculated in two structurally different networks. By using Structural Equation Modeling this paper tested the hypothesis that the time-varying perspective into individual communication behavior will be a strong predictor of emergence in the informal leadership position.

This experiment adds to a growing corpus of research testing the similarity between static and dynamic networks [Soda and Zaheer, 2012, Quintane and Kleinbaum, 2011, Johnson et al., 2012, Zappa and Robins, 2016]. The growing popularity of the electronic communication tools at the workplace contributed to the development of organizational communication research. The structure of the network of email communications is proven to be different from other patterns of contact in the corporate setting with regards to knowledge transfer [Trevino et al., 2000, Grippa et al., 2006], and accuracy of hierarchical system reflection [Rowe et al., 2007, Tashiro et al., 2010]. In line with the previous research I conclude that centrality roles calculated through degree in the network of e-mail exchange corresponds to the degree centrality in self-reported networks [Johnson et al., 2012]. Leadership positions identified through two-years email log data resemble the self-reported social structures in terms of weighted degree centrality modeling the number of existing ties. Results are not consistent among other centrality statistics. Betweenness and closeness centrality show lower correlation correspondence and explain significantly smaller percentage of variance in the informal centrality position. Suggested results are Collectively, obtained results appears consistent with Quintane and Kleinbaum [2011] and show that two networks should be used to answer different questions. The logs of e-mail communications provide a longitudinal perspective on the daily communication patterns of organizational actors, while survey data provides static information about informal status attributed to these actors.

In addition, these findings provide an overview of Structural Equation Modeling as a flexible methodological framework for estimating individual trajectories of change. To my knowledge, the existing literature does not offer a reliable baseline and unique methodological framework for the comparison of the networks constructed from self-reported and electronically registered communications. This study offers the first attempt to use Latent Growth Modeling on social network statistics observed as repeated measurements. Developed models show a good overall fit. Further analysis is required for the validation of results in different organizational contexts.

Using survey as the methodology to gather data about social practices among organizational actors can be labor- and time-consumingand introduce a self-reported bias. Passive methods of gathering the data by parsing electronically registedred communication channels can be less invasive, faster. Such data presents a dynamic view of the interaction however do lacks the context of interaction and requires a higher labor intensity for preprocessing. This research concludes that survey should not be seen as a single reflection of informal organizational structure. Observing long term patterns of previous communications allows to confirm a static picture obtained through the perception of relationship from the survey. In the given example employees with a higher number of citations in corresponding informal networks are expected to be more active in the network of email communications. Understanding which interaction dynamics predicts the outcome existence renders social significance to interactions.

6 Appendices

Table 2: Model selection: Latent Growth Modeling of Weighter Degree Centrality

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Intercept Mean	0.013	0.013	0.013	0.014	0.683
Intercept Variance	x	0.951	0.951	0.973	2.541
Residual Variance	1.024	0.074	0.058	0.058	T1=0.324 T2=0.136 T3=0.054 T4=0.074 T5=0.219
Slope Mean	x	x	x	0.005	0.020
Slope Variance	x	x	0.06	0.06	0.024
Covariance	x	x	x	-0.011	-0.031
Goodness of Fit					
X2(DF)	18(2944.133)	17(350.439)	16(288.829)	14(424.676)	10(245.381)
RMSEA	0.753	0.261	0.244	0.320	0.281
SRMR	0.661	0.019	0.032	0.101	0.105
CFI	0	0.886	0.906	0.859	0.923
Change	x	0.886*	0.02*	-0.047	0.064*
* - change in CFI is significant at the difference of >= 0.01					

Table 3: Model selection: Latent Growth Modeling of Betweenness Centrality

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Intercept Mean	0.012	0.012	0.012	0.012	0.012
Intercept Variance	x	0.776	0.769	0.840	0.855
Residual Variance	1.023	0.248	0.225	0.219	T1=0.222 T2=0.167 T3=0.197 T4=0.309 T5=0.200
Slope Mean	x	x	x	0.000	0.011
Slope Variance	x	x	0.09	0.012	0.024
Covariance	x	x	x	-0.026	-0.027
Goodness of Fit					
X2(DF)	18(1298.052)	17(70.351)	16(55.507)	14(49.611)	10(32.065)
RMSEA	0.498	0.105	0.093	0.094	0.088
SRMR	0.537	0.033	0.051	0.029	0.037
CFI	0.06	0.959	0.969	0.972	0.983
Change	x	0.899*	0.01*	0.03*	0.011*
* - change in CFI is significant at the difference of >= 0.01					

Table 4: Model selection: Latent Growth Modeling of Closeness Centrality

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Intercept Mean	0.029	0.029	0.027	0.012	0.012
Intercept Variance	x	0.292	0.276	0.380	0.367
Residual Variance	0.948	0.656	0.638	0.610	T1=0.629 T2=0.693 T3=0.570 T4=0.582 T5=0.569
Slope Mean	x	x	x	0.09	0.008
Slope Variance	x	x	0.006	0.018	0.019
Covariance	x	x	x	-0.038	-0.035
Goodness of Fit					
X2(DF)	18(217.716)	17(25.119)	16(22.555)	14(17.108)	10(14.330)
RMSEA	0.197	0.041	0.038	0.028	0.039
SRMR	0.224	0.049	0.053	0.039	0.034
CFI	0.029	0.961	0.968	0.985	0.979
Change	x	0.932*	0.007	0.017*	-0.006
* - change in CFI is significant at the difference of >= 0.01					

References

F. Agneessens and R. Wittek. Where do intra-organizational advice relations come from? The role of informal status and social capital in social exchange. *Social Networks*, 34(3):333–345, 2012.

W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the 32nd annual ACM symposium on theory of computing*, pages 171–180. ACM, New York, 2000.

T. J. Allen. *Managing the flow of technology transfer and the dissemination of technological information within the R&D organization.* 1977.

D. G. Ancona, P. S. Goodman, B. S. Lawrence, and M. L. Tushman. Time: A new research lens. *The Academy of Management Review*, 26(4):645–663, 2001.

H. R. Bernard, P. Killworth, and L. Sailer. Summary of research on informant accuracy in network data and the reverse small world problem. *Connections*, 4(2):11–25, 1981.

H. Breitsohl. Beyond ANOVA: An Introduction to Structural Equation Models for Experimental Designs. *Organizational Research Methods*, pages 1–29, 2018.

D. D. Brewer. Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, 22(1):29–43, 2000.

R. Burt. *Structural holes.*, volume 100. Harvard University Press, Cambridge, MA, 1992.

R. S. Burt. The Social Capital of Structural Holes MF Guillén//The New Economic Sociology: Developments in an Emerging Field. In *MF Guillén//he New Economic Sociology: Developments in an Emerging Field.*, pages 148–190. New York: Russell Sage Foundation, 2002.

A. F. de Toni and F. Nonino. The key roles in the informal organization: a network analysis perspective. *The Learning Organization*, 17(1):86–103, 2010.

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*, pages 3–14, 2005.

N. Ducheneaut and V. Bellotti. A study of email work processes in three organizations. *Journal of CSCW*, 2002.

D. R. Farine. When to choose dynamic vs. static social network analysis. *Journal of Animal Ecology*, 87(1):128–138, 2018.

L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

J. Fulk and G. DeSanctis. Electronic Communication and Changing Organizational Forms. *Organization Science*, 6(4):337–349, 1995.

F. Grippa, A. Zilli, R. Laubacher, and P. A. Gloor. E-mail May Not Reflect The Social Network. *SPIE Newsroom*, pages 1–6, 2006.

F. A. Hodis and G. M. Hodis. Latent Growth Modeling for Communication Research: Opportunities and Perspectives. *Annals of the International Communication Association*, 37(1):183–218, 2013.

J. Hox and R. D. Stoel. Multilevel and SEM approaches to growth curve modeling. *Wiley StatsRef: Statistics Reference Online.*, 2014.

E. Ibarra. Network Centrality, Power, and Innovation Involvement : Determinants Of Technical and Administrative Roles. *The Academy of Management Journal*, 36(3):471–501, 1993.

R. Johnson, B. Kovács, and A. Vicsek. A comparison of email networks and off-line social networks: A study of a medium-sized bank. *Social Networks*, 34(4):462–469, 2012.

G. Kannan and W. G. Aulbur. Intellectual capital: Measurement effectiveness. *Journal of Intellectual Capital*, 5(3):389–413, 2004. ISSN 1758468. doi: 10.1108/14691930410550363.

T. Kaye, D. Khatami, D. Metz, and E. Proulx. Quantifying and Comparing Centrality Measures for. *SIAM Undergraduate Research Online*, pages 1–20, 2014.

A. M. Kleinbaum, T. E. Stuart, and M. L. Tushman. *Communication (and Coordination?) in a Modern, Complex Organization.* Harvard Business School, Boston, MA, 2008.

A. M. Kleinbaum, T. Stuart, and M. Tushman. Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organization Science*, 24(5):1316–1336, 2013.

G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757): 88–90, 2006.

D. Krachardt and J. Hanson. Informal networks: the company behind the chart. *Harvard Business Review*, 71(4):104–111, 1993.

G. Krings, M. Karsai, S. Bernharsson, V. D. Blondel, and J. Saramäki. Effects of time window size and placement on the structure of aggregated networks. *Biomedical Engineering*, page 19, 2012.

R. Lambiotte, V. D. Blondel, C. D. Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, 2018.

P. V. Marsden. Network data and measurement. *Annual review of sociology*, 16(1):435–463, 1990.

P. V. Marsden. Recent developments in network measurement. In *In P. J. Carrington, J. Scott & S. Wasserman (Eds.), Models and methods in social network analysis*, pages 8–30. 2005.

J. Melorose, R. Perroy, and S. Careas. Exploration of Communication Networks from the Enron Email Corpus. *Statewide Agricultural Land Use Baseline 2015*, 1:1–12, 2015.

P. R. Monge and N. S. Contractor. *Theories of communication networks.* Oxford University Press, USA., 2003.

A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi. On the Structural Properties of Massive Telecom Call Graphs : Findings and Implications. In *Proceedings of the 15th ACM international conference on information and knowledge management.*, pages 435–444. ACM, New York, 2006. ISBN 1595934332.

J. A. Odero and W. Otuya. Critical Review of Literature on Knowledge Management Strategy and Organizational Performance. *International Journal of Management and Commerce Innovations*, 5(2):741–748, 2017.

J. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

E. Quintane and A. M. Kleinbaum. Matter Over Mind? E-mail Data and the Measurement of Social Networks. *Connections*, 31(1):22–46, 2011.

R. Rowe, S. J. Stolfo, and R. Rowe. Segmentation and Automated Social Hierarchy Detection through Email Network Analysis Automated Social Hierarchy Detection through Email Network Analysis. (January), 2007.

M. Seshadri, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec. Mobile Call Graphs : Beyond Power-Law and Lognormal Distributions. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 596–604. ACM, New York, 2008.

B. H. Sims, N. Sinitsyn, and S. J. Eidenbenz. Hierarchical and matrix structures in a large organizational email network: visualization and modeling approaches. In *Social Network Analysis-Community Detection and Evolution*, pages 27–43. Springer, Cham, 2014.

G. Soda and A. Zaheer. A network perspective on organizational architecture: Performance effects of the interplay of formal and informal organization. *Strategic Management Journal*, 33(6):751–771, 2012.

H. Tashiro, J. Mori, N. Fujii, and K. Matsushima. Email Network Analysis for Organizational Management. (June 2009):958–963, 2010.

L. K. Trevino, J. Webster, and E. W. Stein. Making Connections : Complementary Influences on Communication Media Choices , Attitudes, and Use. *Organization Science*, 11(2):163–182, 2000.

J. Tyler, D. Wilkinson, and B. Huberman. Email as spectroscopy: automated discovery of community structure within organization. *The Information Society*, pages(2):81–96, 2003.

S. Wasserman and K. Faust. *Social network analysis: Methods and applications.* Cambridge university press, (vol. 8) edition, 1994.

E. Yan and Y. Ding. Applying Centrality Measures to Impact Analysis. *Journal of the American Society for Information Science and Technology*, 60(10):2107–2118, 2009.

P. Zappa and G. Robins. Organizational learning across multi-level networks. *Social Networks*, 44: 295–306, 2016.