Aalto University

School of Science

Degree Programme in Information Networks

Marko Rossi

A Comparison of Intra-Organizational Networks of Project Participation and Discussions

Master's Thesis

Espoo, 21/11/2014

Supervisor: Professor Eila Järvenpää

Instructor: D.Sc. (Tech.) Anssi Smedlund

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Marko Rossi

Title of thesis A comparison of intra-organizational networks of project participation and discussions

Degree programme Degree Programme in Information Networks

Thesis supervisor Eila Järvenpää **Code of professorship** TU-53

Department Department of Industrial Engineering and Management

Thesis advisor(s) Anssi Smedlund

Date 21.11.2014 **Number of pages** 88+5 **Language** English

Abstract

Communication in an organization as well as how the work is really done rarely obey the formal organization chart. The connections people have within the organization form informal networks that are usually difficult to see clearly without the help of social network analysis. Social networks have been traditionally studied by using surveys. Firms also have different kinds of databases containing records of employees' participation in different events, such as projects, meetings or courses. These kinds of databases offer an alternative way to collect information on social networks in the organization.

This thesis compares two self-reported communication networks and a network inferred from the project time tracking database. The two self-reported networks describe whom employees discussed on routine work related matters with, and the second network describes whom they discussed ideas with. The network data for the self-reported networks and the project participation network were collected from an architect's office in Northern Europe in 2007. The research method used is a social network analysis focusing only on the structural properties of the networks without attribute data of the employees.

A hypothesis was formed that the network of project participation has similarities with the network of discussing routine work matters. The network of discussing ideas was expected to be different to these other two networks. The results show that these three networks were all different, but the network of project participation had more similarities with the network of discussing routine work than with the network of discussing ideas. Participation to projects and discussion on routine work show more cohesion whereas ideas form sparser and less cohesive network structure. The thesis shows that the communication structures of routine work and discussing ideas are structurally different. It also demonstrates that using social network analysis can be a helpful tool to understand the social structures in organizations.

Keywords Social networks, Social network analysis, SNA, Routines, Ideas

Aalto-yliopisto
Perustieteiden
korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Diplomityön tiivistelmä

Tekijä Marko Rossi

Työn nimi Vertailu projekteihin osallistumisen ja keskustelun verkostoista organisaatiossa

Koulutusohjelma Informaatioverkostot

Valvoja Eila Järvenpää **Professuurikoodi** T-53

Laitos Tuotantotalouden laitos

Työn ohjaaja(t) Anssi Smedlund

Päivämäärä 21.11.2014 **Sivumäärä** 88+5 **Kieli** englanti

Tiivistelmä

Kommunikaatio ja kuinka työ organisaatiossa tehdään harvoin noudattavat virallista organisaatiokaaviota. Yhteistyö ja kommunikointi organisaation jäsenten välillä muodostavat verkoston, joka voi olla vaikea hahmottaa ilman sosiaalisten verkostojen analyysin menetelmiä. Sosiaalisia verkostoja on perinteisesti tutkittu kyselyjen avulla. Yrityksillä on myös erilaisia tietokantoja, jotka sisältävät tietoja työntekijöiden osallistumisista tapahtumiin, kuten erilaisiin projekteihin, kokouksiin tai koulutuksiin. Nämä tietokannat tarjoavat vaihtoehdon tietojen keräämiseen sosiaalisista verkostoista.

Tässä työssä verrataan kahta kyselyihin pohjautuvaa verkostoa ja verkostoa, joka on päätelty työntekijöiden projekteille raportoiduista työtunneista. Kyselyihin pohjautuvat verkostot kuvaavat, keiden kanssa organisaation jäsenet keskustelivat rutiinityöhön liittyvistä asioista ja keiden kanssa he keskustelivat ideoista. Verkostoaineisto on kerätty Pohjois-Euroopassa toimivasta arkkitehtitoimistosta vuonna 2007. Tutkimusmenetelmä on sosiaalisten verkostojen analyysi, ja vertailun kohteena ovat rakenteelliset erot verkostoissa. Taustamuuttujien vaikutusta ei huomioida.

Verkostojen eroista muodostettiin kirjallisuuteen pohjautuen hypoteesi, jonka mukaan projekteista muodostunut sosiaalinen verkosto muistuttaa enemmän rutiinityön keskustelun verkostoa kuin ideoiden keskustelun verkostoa. Analyysin tuloksena huomattiin, että kaikki kolme verkostoa olivat erilaisia. Projekteista muodostettu sosiaalinen verkosto muistutti hypoteesin mukaisesti enemmän rutiinityön keskustelun verkostoa kuin ideoiden keskustelun verkostoa. Osallistuminen projekteihin ja keskustelu rutiinityöstä muodostivat tiiviimpiä ryhmiä kuin ideoista keskustelu, ja keskustelu ideoista muodosti näitä verkostoja harvemman verkoston. Työ osoittaa, että sosiaalinen vuorovaikutus liittyen rutiinityöhön ja ideoihin ovat rakenteellisesti erilaisia. Lisäksi työ näyttää, että sosiaalisten verkostojen analyysi on käyttökelpoinen väline kommunikaatiorakenteiden ymmärtämiseen.

Avainsanat Sosiaaliset verkostot, sosiaalisten verkostojen analyysi, rutiinit, ideat

Acknowledgements

I would like to thank D.Sc. (Tech.) Anssi Smedlund and Professor Eila Järvenpää for the interesting subject for the thesis and for insightful comments, and for letting me borrow books way too long time.

And I also thank all the other usual suspects as well. Especially the dear librarian there.

And you, the reader, thank you too that you are reading this.

Espoo, November 2014

Marko Rossi

Table of Contents

List of Appendices

List of Figures

List of Tables

1 Introduction

1.1 Background

The term network can have different meaning in different disciplines. In the management literature, it is often used liberally to describe what organizations have to become in order to be dynamic and competitive. In social science, networks are a way of seeing social systems in terms of relationships between actors. Actors can be individuals, teams, organizations or even nations.

Social networks have been traditionally studied by using surveys. Rapid growth of social media sites and communication tools have provided researcher new ways of collecting network data directly with database queries. Regardless of what media are used for communication, people usually have some mutual reason to have a connection. They might have met before somewhere or discussed because of shared interest in something.

Informal networks consist of employees and their ties to other employees and define whom they discuss with and ask advice, opinions and ideas. Ties can form, for example, by participating in the same events. Informal networks have important consequences for organizations, because how the work really gets done has usually more in common with informal networks than with the formal organizational chart. The research question in this thesis is related to these kinds of informal networks. The focus of this thesis is on comparing self-reported networks with the network that emerges from working on the same projects.

In many professional service firms, work is organized around projects and projects connect actors together. There is plenty of research on networks formed by individuals participating to events, but not from the perspective of comparing self-reported communication networks to network formed from project participation. There are published research articles in social network literature related to the social networks formed by participating to task groups (e.g. Quintane et al. 2013). Several studies have compared self-reported and observed networks (e.g. Kilduff et al. 2008). In addition, social networks of email-communication have been studied (e.g. Johnson et al. 2012; Quintane & Kleinbaum 2011). This thesis aims to fill this gap partly by comparing project co-occurrence networks and communication networks of networks as they look like in a professional service firm.

The research problem in this thesis is to compare two self-reported communication networks and a network derived from the project work hour reporting database. The network data for the self-reported networks and the project participation network was collected from an architect's office in Northern Europe in 2007.

The third network will be constructed from the work hour data. The project working hour data contain records on which projects each employee worked on and how many hours they reported work for each of these projects. This network shows which actors participated in the same projects and how many hours each pair of actors had possibly worked together.

The research problem is to find how these networks are different or if they are similar in some way. The goal of this study is to find answers to the research question using social network analysis while focusing on structural differences and similarities. Structural here means, that the interest is in the structure and patterns of relationships, instead of the attributes of the members of the networks or the social processes forming these networks.

Based on the research problem, the following research question was set to guide the study:

What kind of structural differences and similarities do the networks of project participation and discussing routine work and ideas have?

The theoretical objective of the study is to review the research literature of networks in organizations, organizational routines and idea generation and creativity. In addition, necessary methodology and foundations of social network analysis are reviewed. The next objective is to formulate hypotheses of the expected results based on the theoretical review. The third objective is to answer the research question by performing a multilevel analysis of each of the networks: from the whole network level to the individual actor level. Multilevel analysis is hoped to either find support of falsify the formulated research hypotheses. The Fourth objective is to evaluate the results against previous literature and to see if the findings reveal possibilities for future research.

1.2 Scope of the study and research methods

Social network analysis has traditionally focused on structural properties of the networks more than attributes of the individuals (Brass et al. 2004). Empirical part in this thesis follows this practice and focuses on analyzing the network structure. Various attributes, although important, are not considered in structural analysis. It could be possible for example to estimate a statistical model to explain the relationship between attributes and structure, but this is left for further research.

The analysis methods were limited to the most well-known analytical tools that are already implemented in most social network analysis packages. The goal is to use several different methods to form a reasonably complete description of the structural properties of the networks.

The methodology of social network analysis is vast and constantly developing. Stochastic models of the network structure such as exponential random graph models are left outside this study. Although they are becoming common and are used extensively in research literature, model specification and fitting a convergent, and theoretically sound model can be computationally demanding. In addition, these methods are best used to estimate a theory based model consisting of explanatory variables, such as various attributes of the individuals while controlling for the feedback of endogenous variables.

The research methods of this thesis are literature review for the theoretical part and social network analysis for the empirical part. Attribute-based statistical methods such as various regression models are not used.

The empirical part of this work is based on a dataset made available for the use in this thesis by Anssi Smedlund. The data were collected in 2007 in a study of routine and non-routine communication networks in a professional service firm (Smedlund & Choi 2009). The case company is an architect's office in northern Europe whose design and planning work is highly knowledge-intensive. The network data consist of persons located in the main office and the connections they reported having to others. The boundaries of the three network studied were limited to within the organization.

1.3 Structure of the thesis

The theoretical part of this thesis starts from chapter 2, which gives a theoretical review of the network perspective to organizations and introduces theories of organizational routines and creativity and communication of ideas. The theoretical ideas presented here are used as a theoretical base to construct hypotheses of the network structure of the three different networks studied. These are presented in chapter 3.

Chapter 4 continues by introducing research methods and the data used in the empirical part. It also explains how the data was prepared and provides an overview of the data.

Chapter 5 contains the results of network analysis. Because the results of the analysis are presented in many tables and figures, these are presented together with each analytical method used. The results are then combined in chapter 6.

Chapter 6 discusses the results of the empirical part, reflects them back to the ideas presented in the theoretical part and discusses the sources of errors and the limitations of the study. There are also suggestions for further study. Chapter 7 concludes the thesis.

2 Theoretical review

2.1 Social network perspective to organizations

Social network perspective is a distinct research perspective within social and behavioral sciences based on the assumption of the importance of relationships among interacting actors. The social environment is expressed as a structure of relationships with patterns and regularities. The unit of analysis is not an individual, but the whole collection of individuals and the linkages among them. Actors are seen to be embedded within networks of interconnected relationships that both provide opportunities for and constrain the actors. (Kilduff & Brass 2010; Borgatti et al. 2013; Wasserman & Faust 1994)

Traditional social science research focuses typically on attributes of autonomous individuals, the associations between attributes and their predictive power. Social network perspective sees these characteristics arising out of structural processes. In practice, it means that relational ties are seen as the primary source of information and attributes come second in importance. The goal is to understand properties of the social structure and how these can explain many social phenomena. (Wasserman & Faust 1994)

One of the advantages of the social networks perspective is that it allows to examine organizational phenomena at different levels. From the macro to the micro level, social network research has been used to study topics such as inter-firm networks, alliances, network governance, employee performance, power, attitude similarity and innovation and creativity. (Kilduff & Brass 2010)

Kilduff and Brass reviewed organizational social network literature and recognized four leading ideas that are on the center of much of the research program and which act as a base for new theories. The most fundamental idea where the whole field is based on is the importance of relations between actors. The second key idea is that individuals and organizations are embedded in a social structure. Third, there is recognizable patterning in social relations and finally that there is utility value in social ties. (Kilduff & Brass 2010)

2.1.1 Importance of social relations and social structure

Although social network analysis is a multidisciplinary field, historical accounts tend to agree that it has roots in social psychology and Gestalt theory. Jacob L. Moreno explored how social relations affect psychological well-being in a book on sociometry, "Who Shall Survive?" (Moreno 1934). Researchers in social anthropology field started focusing independently in social structure in 1940s. The field of sociology has had an interest in social structure even as a field it did not adopt social network analysis early. Sociological thinkers, such as Simmel, Durkheim and Weber have been influencing the social network view in other fields much earlier. (Prell 2012)

Network research has tended to have a Durkheimian approach to network research that emphasizes relationships over actor attributes, but attributes of actors also have attracted attention (Kilduff & Brass 2010). For example, research on self-monitoring personality character has provided some evidence that what kind of network positions people occupy depends on how high or low they are on self-monitoring orientation (Oh & Kilduff 2008). Strategy research has focused on firm-specific characteristics and distinctiveness of firm's capabilities as a source of competitive advantage, which is sometimes called the resource-based view of strategy (Teece et al. 1997). Network research has later combined these views by focusing on the properties of the actors affecting the focal firm's performance. (Kilduff & Brass 2010; Borgatti & Foster 2003)

The network approach is built on the assumption that there are persistent patterns of connectivity and absence of social ties under social relationships that can explain social outcomes. The overall social structure can be characterized by indicators such as clustering, connectivity and centralization. (Kilduff & Brass 2010; Brass et al. 2004). Relationships are visible on multiple levels and tied together on multiple levels. Organizations are connected to each other through people and people can be connected to each other through organizational affiliations.

Often different kinds of groups emerge in organizations, structuring the social network. Groups are formed, for example, through homophily, a preference for similar people or active recruitment of friends or people with similar background. These kinds of groups tend to be denser with relationships than the space between different groups. Dense networks have redundancy in the different paths along which information and influence can flow. Dense networks also tend to have norms concerning the proper behavior. (Kilduff & Brass 2010)

2.1.2 Embeddedness

Another distinctive idea in social networks research is the idea of embeddedness, that the human action and economic transactions occur within the larger context of social relationships. The idea is often credited to Karl Polanyi, who proposed that in market societies, economic action is embedded in formal market structures and social structures, and in nonmarket societies economy is submerged in social structures only. (Kilduff & Brass 2010, p.323).

Mark Granovetter developed the idea of embeddedness further, emphasizing the role of social networks in guiding all kinds of economic transactions. According to Granovetter (1983), economic research has tended to go into under-socialized and over-socialized directions. Under-socialized classic economics sees people acting atomistically, paying attention only to own self-interest and maximization of economic utility. This is especially the view taken by neoclassical economic, which sees relationships being impersonal ties that occur at dyadic level without the larger network structure. In an over-socialized view, actors are seen agency free, regulated by inherited norms and values, acquired through socialization with people around them. (Granovetter 1983; Granovetter 1985)

Granovetter's view was a middle of the road perspective: rational economic actions is embedded in social relationships that constrain and structure the transactions. Social interaction produces trust that helps to reduce uncertainty in transactions and creates new economic opportunities. The middle of the road view argues that actors are not solely interested in a narrow pursuit of economic profit. Attention have been paid on additional effects of other aspects in transactions, such as the pursuit of enrichment of relationships through trust and reciprocity. (Granovetter 1983; Granovetter 1985; Powell 1990)

After Granovetter's discussion of embeddedness, social networks have been used to explain various economic actions taken by individuals and firms. Padgett and Ansell (1993) studied elite networks of Renaissance time Florence. Their dataset included relation information of marriage and business ties. Powell (1990) expanded the concept of embeddedness further, seeing it as organizing logic in inter-organizational collaboration. Uzzi (1996) refined the embeddedness principle with tendency for actors to cultivate long term relationships and repeating transactions.

One type of criticism of social network research concerns it failing to take into account human agency. People take intentional action, which also establishes social networks where they are embedded in. Traditionally, social network research has focused mostly on actors being at the right place in social network while ignoring the psychological processes leading actors to that position in the first place. Actor's cognition also has a role in network formation. According to Balance Theory (Heider 1958), people expect their friendship relations to exhibit reciprocity, and they suffer emotional tension if they perceive their friends fail to reciprocate friendship to each other. As is said, you are known by the company you are with, and it affects how people relate to you. (Kilduff & Brass 2010)

2.1.3 Network connections and social capital

Networks are believed to provide opportunities and set constraints that affect the outcomes and the possibilities individuals and group have in organizations. Probably the biggest area of the research of the utility of network connections is the concept of social capital.

The concept of social capital has been around in the literature starting from the early 20s when it was introduced by Lyda Judson Hanifan (1916) but appeared in organizational studies in 1980s (Coleman 1988). Putnam (1995) described social capital meaning "features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit". In research literature, social capital has been seen to consists of social networks, norms and beliefs such as shared vision and meanings (Smedlund 2008). Nahapiet and Ghoshal (1998) recognized three dimensions in social capital: a cognitive dimension for the formation of norms, a structural component forming the context and relational dimension forming motivation.

In management literature, social capital has often been seen from a rent-seeking point of view while sociologist has focused on how individuals benefit from it (Blyler & Coff 2003). In organizational studies, different viewpoints have been taken in a similar manner. Social capital benefits individuals (Granovetter 1973; Burt 2002) and community as a whole (Coleman 1988; Putnam 2000).

According to Mark Granovetter (1973), acquaintances outside one's nearest social group facilitate flow in information that is not present in the group of the closest friends. Granovetter called the connections to acquaintances *weak ties* and argued that weak ties are important in connecting otherwise separated components in a social network together. There is no easy way to define which ties are weak ties and which are strong ties. It is a function of length of knowing someone, frequency of interaction or sentiment of closeness a pair of actors feels. The strongest ties are those ties with most of these characteristics. Weak ties might not be reciprocal and involve less interaction. Discussions through weak ties might be only related to work matters without many other social exchanges. Because weak ties involve less affection, the parties might be very different to each other and be members of different social circles. For this reason, weaker ties are likely to bridge people with different perspectives and interests and new approaches together (Granovetter 1973).

A major theory combining social capital to social network structure is structural-hole theory introduced by Ronald Burt (1992). The idea of structural holes is built on the notion that people tend to form cohesive groups. How people behave, what kind of opinions they form and what information they have is more homogeneous within group than between groups, because people tend to focus more on activities inside their own group. This creates holes to information flow between groups (Burt 2004).

The structure of actor's ego-network or specifically the lack of ties among actor's alters provides opportunities for bridging connections between parts of the network. Bridging connects separated network parts together permitting new information to flow between groups. People in this position have early access to new information arriving from multiple parts of the network whereas agents within dense network components have redundancy in their information flow. People at structural holes also have chances of creating novel ideas by combining knowledge from multiple sources, and they act as gatekeepers providing them with a source of power. One well known property of social networks is the short path lengths: there are many alternative routes from one individual to another. This suggests that structural holes are not very common in social networks. (Borgatti & Foster 2003; Kilduff & Brass 2010; Burt 2002; Burt 2004)

There is an opposing view to social capital which pays attention to dense ego-networks. Dense components provide trust, norms and reciprocity and support of the nearby people. Cohesion affects individual's social identity, and the clear normative order in a cohesive group provides individual possibility to optimize her performance. Redundancies in ties add common understanding. Structural holes theory values new knowledge over existing accumulated knowledge base (Obstfeld 2005). Dense networks are good environments to share complex, tacit information which is often crucial for knowledge intensive firms and their innovativeness. In addition, getting help from a sparse network is not easy. (Uzzi 1997; Hansen 1999)

Dense network might pose a problem for idea generation as information in cohesive clusters tends to be redundant. Individuals might feel good with strong relationships, but this does not increase exposure to relevant new ideas (Perry-Smith & Shalley 2003). Good ideas need to be implemented in some point as well. Dense networks reduce the obstacles to implementing the good ideas and help to coordinate action. Sparse networks, on the other hand, are not that easily coordinated as many people are disconnected (Coleman 1988; Granovetter 2005)

Both structures seem to provide advantages, but overall the current position seem to be that structural holes offer opportunities for new information, innovative ideas and competitive brokerage. Cohesive networks offer collaboration, better transferring of knowledge, implementation of ideas and better learning opportunities for complex knowledge. Both, the individual and the society as a whole benefit from social capital. (Kilduff & Tsai 2003)

Apart from the benefits, social network researchers agree that cohesion is constraining. Cultural context can change the effect of cohesion. There is also research on the dark side of social capital, where social ties imprison actors to harmful position or encourage undesirable behavior (Borgatti & Foster 2003, p.994)

2.2 Features of social networks

Ties can form between people for many reasons. People may be connected because of shared interests, similar personal characteristics, because they participated the same events or because they are working on the same project. Geographical proximity is one important explaining factor in people's connections (e.g. Feld & Carter 1998).

Explanation why network ties form in the first place are offered from psychology: Deeper human needs to feel safe and to stay near familiar people, need to reach out and need to seek higher status in social hierarchy guide how people make connections with others. Interaction with other leads to sentiments which further affect the future interactions (Kadushin 2011, p.110).

A pair of connected actors is called a *dyad,* and a group of three people are called a *triad*. Dyads and triads are the basic building blocks in social networks. Mutuality, tendency of shared friends becoming friends, and popular people getting even more popular are basic processes visible in social networks. Triads already add significant complexity to the network structure as there can be 16 different variations of how the ties in a triad can be formed. Heider's Balance theory states that only certain types of triads are stable: people tend to maintain consistency in their liking and disliking patterns of other people (Munroe 2007). Dyads and triads form a basis for many social network analysis methods.

People who are near each other tend to share similar values, social statuses and other characteristics. A well-known observation is that people are more likely to have social ties with those similar to themselves on important attributes such as the gender, race, education and age. This observation is known as homophily (McPherson et al. 2001). Direction of causality between homophily and network ties is not clear however: people can be similar to each other because they spend time together or are near each other, or they spend time together and are near each other because they are similar. This demonstrates that social networks are dynamic and that there are self-enforcing feedback loops changing them.

Even though social network analysis focuses mostly on structure, characteristics of individuals are important for how ties between them form. For example, whom people discuss with their routine work or seek advice from probably depends on the attributes of the people, such as work experience, job position and language. Many features of social networks also apply to networks of whole organizations. Similar firms tend to be located near each other (for example, Hollywood and Silicon Valley) and ties form easier between organizations sharing some similarities.

Social networks typically show cohesive clusters that are connected to other clusters with bridging actors (Burt 2004). To some extent, social networks show small world properties (Kilduff et al. 2008). A small-world network is a network with short overall path lengths and a large amount of clustering (Watts & Strogatz 1998). This kind of property might follow from homophily and weak ties (Jackson 2010). Social networks might not be purely small-world networks, but they often have a high number of clusters and hubs, highly connected actors, and distances between any two actors are short (Kilduff et al. 2008).

2.3 Social networks in organizations

Organizations are social arrangements aiming to achieve controlled performance in pursuit of specified goals through cooperation of individuals. Formal organizations have been designed to facilitate standard operations and to handle easily anticipated events (Buchanan & Huczynski 2004, p.5). Organizations utilize a chain of authority to decide what to do, to acquire needed resources and to get people accomplish what is required to reach the specified goals.

Formal organization only partly reflects how work is done. Much of the work and exchange of knowledge is done outside the formal organization chart in informal networks (Krackhardt & Hanson 1993). Informal networks can form similarly to friendship networks, but they can also form from pragmatic reasons. Employees can skip the formal organization to work the system to suit their needs. In the knowledge-intensive work employees must solve complex problems within short time frames. For this reason, professional workers often maintain their own contacts and sources of information. Formal organizations also offer political reasons for people choose whom to interact with. (Cross et al. 2002).

Nature and existence of optimal communication network structure has been a source of disagreement in social capital literature. Uzzi (1997) analyzed embeddedness in the network of different firms in New York's garment industry. He showed that being embedded into the network of suppliers brought positive returns, but only up to a certain threshold. Being embedded beyond certain threshold level can act adversely by reducing adaptive capacity. Arm-length's relationships offer flexibility and don't cause a lock-in to existing relations. Optimal networks in the garment industry were not composed of either fully embedded or arm-length ties but something between the two. Hansen (2001) analyzed data from 67 new product development teams and found that the optimal structure of the communication network is contingent on the task type teams were working on. Teams that were working on explorative tasks that departed from the existing expertise of the network performed better if the team had many strong ties. Teams that exploit existing knowledge were slower to complete the task if they had strong ties. Strong ties are costly to maintain and might not be needed in every task (Hansen 2001). The best structure seems to depend on the tasks being completed, and there is no single optimum for network structure, but many (Smedlund 2008).

Individuals in a network may have an accurate view of their immediate social circles, but inaccurate view of the network beyond their direct contacts. Managers' perceptions of the social networks in their organization can differ from what can be seen in results from social network analyses conducted in the organization. (Cross & Parker 2004)

The number of ties in a network grows quickly when the network grows larger. Even relatively small organizational networks require keeping track of hundreds of possible ties. Complexity of social networks is likely to cause a cognitive challenge to individuals. Kilduff et al (2008) suggest that the mental map individuals have of the network is based on the small world principles: The structure of the friendship relations is seen to exhibit clustering, centralization, small path lengths and good connectivity. The mental schemas actors have of the networks don't necessarily coincide with reality. Schemas simplify the structure and exaggerate the properties perceived in it, but at the same time help the individual to recall the structure of the network (Kilduff et al. 2008). This has consequences for social network analysis: the results of network surveys are distorted by respondents' cognition. For example, groups based on friendship can be seen more united than groups based on tasks or projects (Lickel et al. 2000). Friends are mentally classified to friendship-category, and respondents might fill in some of the blanks to create more clear friendship clusters than reality presents. (Kilduff et al. 2008)

2.4 Organizational routines

Routines in organizations have got attention in fields of evolutionary economics and strategic management. In everyday parlance, routines refer to repeated sequences of behavior. In organizational settings, routines may be considered as an organizational analog of individual habits. Constant change of the economy suggests that there is a Darwinian kind of evolutionary process associated. It has been proposed that how firms adapt and behave can be understood based on the routines they have (Nelson and Winter, 1982).

There have been at least three different kinds of interpretations of routines. Routines have been characterized as behavioral regularities, cognitive regularities or propensities and as stored behavioral capacities with knowledge and memory (Becker 2004).

Routines are relatively enduring behavioral patterns in organizations, including rule like structures, forms, procedures, production techniques and policies (Hodgson 2008). Routines have been even argued to be in a similar role to genes in biological evolutionary theory (Nelson & Winter 1982). This analogy refers to the idea that routines carry information in a relatively durable way and are generative rule-like structures and related with procedural memory in organization (Becker 2004).

Routines are built on habituated individuals in the social structure, but routines are not just a collection of habits shared by individuals. For example, individual habits can deviate from organizational routines, sometimes harmfully as found in relation to aviation safety (Weick 1990). Routines also are not just behavior as routines can include capabilities or potential to change or cease behavior. Routines are collective phenomena and can also be regarded as a collective analogy to individual skills (Nelson & Winter 1982).

Local groups and communities of practice provide different contexts for routines to emerge. Routines are embedded in social structures in a specific context and in specific relations. They are also path dependent on history: they can get stuck to the path they were developed in and the starting point matters. For these reasons, routines are difficult to replicate or to transfer elsewhere from their original context. Because of the dependencies to context and history, there are no universal best practices applicable to all organizations but local best practices (Becker 2004; Amit & Belcourt 1999).

Routines are consequential for organizations in many ways. They offer coordination and control to the firm; they add trust, build predictability in the work and provide automation for recurring tasks. This way routines reduce uncertainty and increase stability. Finally, routines act as important storage of organizational tacit knowledge, and they maintain the artifacts of codified knowledge (Becker 2004).

Routinization of activity stores operational knowledge into organizations (Nelson & Winter 1982, p.99). It is difficult to get big-picture of the knowledge that is embedded in the social structure. Routines are distributed across the organization, and they vary between different groups and parts of organization (Becker 2004).

2.5 Communication in organizations

Effectiveness of communication is essential for organization's performance. Knowledge intensive work is rarely done alone, and most of the managerial tasks involve communication. Traditional organization design however is effective at inhibiting communication because of levels of hierarchies, status differences and separation by work division and departments (Buchanan & Huczynski 2004, p.179).

Communication does not have to follow formal organizational hierarchy, but it does not occur without restrictions either. Research literature reveals many important factors affecting likelihood of communication in organizations. These factors are related to attributes of the both parties of communication and attributes of the relationship and the organizational culture (Cross & Borgatti 2004). Communication is easiest when people share some similarities and is likely to occur in homophilous relationships (McPherson et al. 2001). Being in the same place with an opportunity to communicate and psychological obligation for communication is also recognized as being important requirement and is referred to as organizational proximity (Monge et al. 1985). Even if communication does not follow the formal chart, position in the formal organizational structure is a factor affecting communication (Stevenson 1990).

Individuals have socially learned and confirmed expectations on each other and on proper behavior in organizations. They also hold assumptions about other people's motives, intentions and prospective actions. There has been interest in trust in organizations and its effect to organization's performance. Trust turns out to be important in allowing people to share valuable information to each other (Kramar 1999).

From social network perspective, the interest is mostly in the nature of social ties connecting individuals to the social structure. Weak ties are important in their ability to bring new information from outside to one's own social circles. On the other hand, people whom one has strong ties with are more readily available (Granovetter 1983).

Research on communities of practice has produced knowledge on how people on cohesive subgroups communicate. A community of practice is a tightly knit group of people who have shared interest in something they do, and learn how to do it better as they interact regularly. It is formed either specifically in order to gain knowledge in some field or has emerged naturally by member's common interest in particular area (Wenger 2000; Lave & Wenger 1991). Its members know each other and typically meet face-to-face. Communities of practice are characterized by strong interpersonal ties and norms of reciprocity in connections. The members of the community of practice learn from each other by discussing their work, by sharing tips and best practices and by providing support to each other (Lave & Wenger 1991)

Cohesion and face-to-face discussion provides trust, creates expectations and obligations for reciprocity and encourages knowledge sharing (Wasko & Faraj 2005). The strong ties in cohesive groups are important for sharing tacit and complex knowledge (Hansen 1999). Groups tend to form in organizations based on similarity, familiarity and proximity. When people are similar, communication is easier and predictable (Perry-Smith & Shalley 2003). Community of practice literature shows that tenure difference affects information sharing. Employees with younger career would seek information from more experienced employees (Wasko & Faraj 2005).

Borgatti and Cross (2003) studied how people in organizations intentionally seek information from colleagues. They proposed that there are relational characteristics that facilitate and predict information seeking from the organization. Specifically, information is sought from people whom actors know to have expertise and estimate the level of expertise to be valuable. The person should be easily accessible and seeking information from that person should not be too costly, for example, by admitting own ignorance of a given subject. (Borgatti & Cross 2003; Cross & Borgatti 2004)

Generally, role of face-to-face communication seem amplified, when the information needed involves technical knowledge (Allen 1977). Globalization of business has brought increasingly multicultural work environments. Face-to-face communication not only requires language skills, but also sensitivity to the norms and expectations of other cultures (Hofstede 1991).

2.6 Ideas in organizations

Organizational routines might not be enough when the firm has to develop new products or its own operations or solve complex problems. Most firms need constantly to innovate and improve in order to sustain profitability. Innovativeness requires combining new ideas and existing knowledge in a novel way and new ideas in turn require creativity. Innovations require creativity, but creativity doesn't require that ideas must become innovations. There have been large amounts of research on creativity and innovation and in multiple levels covering organizations, teams and individuals. Innovation and creativity have been studied separately, where the creativity part focuses on the generation of ideas, and innovation involves much more complex topics including the implementation and business value of ideas. Creativity has been often researched for individual level factors such as personality, job satisfaction and motivation. (Anderson et al. 2014; Parzefall et al. 2008; Axtell et al. 2000)

According to Koestler (1989), ideas emerge from a new intersection of already existing knowledge or assumptions. Ideas come up as solutions for a problem or a puzzle. First the mind soaks up information regarding the problem. There is an incubation time when the unconscious mind works the problem, and the solution may show up suddenly later (Koestler 1989). In addition to creativity, generation of workable ideas requires knowledge and learning. Much of the knowledge work in organizations involves tacit knowledge located in the social structure. Social networks are thus in important role. Ideas are partly born from the interaction with others in the social network (Spender 1996).

From the organizational networks point of view, there are at least two possible views on ideas following the debate on cohesion and structural holes in social capital literature. Because ideas require combining existing individual and shared knowledge with new knowledge, cohesive groups such as communities of practice might be important. These facilitate knowledge sharing and learning and allow new combinations of information. Implementation of new ideas might need social support and acceptance from managers and co-workers. Sharing ideas might have a political need to gather enough support for implementation. (Ohly et al. 2010; Axtell et al. 2000)

On the other hand, weak ties have been seen important in exposure to fresh ideas outside one's own social circles. Boundary-spanning individuals have an important role to let fresh information flow into more cohesive groups. For example, Burt (2004) analyzed several hundred managers and found support for the hypothesis that managers near structural holes are more likely to come up with good ideas. Perry-Smith (2006) found that in a network formed by research lab researchers, weak ties were more important for new ideas than strong ties.

Perry-Smith (2003) proposed that not only will weak ties facilitate creativity and generation of ideas at work tasks better than strong ties, but that strong ties will constrain creativity. Strong ties tend to connect similar individuals, and with time, lead to social pressures to conform to group's standards without much room for autonomy. Perry-Smith further suggested that more weak ties are better for ideas, but because there is a limited amount of time for all the contacts, too many weak connections start to hinder creativity. For example, a busy manager connected with many weak ties could be too busy to come up with good ideas. According to Csikszentmihalyi (1996), creativity needs a certain amount of focus, attention and mental energy.

Actors who show up as peripheral in network analysis might be important to idea generation. As network analysis always contains boundary setting decision, some actors will look peripheral in the network even though some of them might be central in some other network and act as bridges between different networks. Peripheral actors are not tightly embedded and can notice new ideas and approaches without worries of breaking norms of the group. Peripheral actors might see themselves as members of different social systems and the social pressure they experience might come from elsewhere. (Perry-Smith & Shalley 2003)

3 Research question and hypotheses

Chapter 2 presented a theoretical review of social networks in organizations and reviewed the concepts of routines, ideas and creativity in an organization. This section combines the theoretical findings to formulate hypotheses from the research question of how the network structures of discussion of routine work and discussion of ideas differ.

Much of the complex knowledge in professional service firms tend to be tacit because diverse customer cases limits the usefulness of codified knowledge (Hansen et al. 1999). Tacit knowledge is transferred through regular personal contacts in an environment characterized by trust (Nonaka & Tackeuchi 1995). Smedlund (2008) suggests that social networks are in important role in facilitating the distribution of knowledge within the organization.

The employees in the case company worked mainly on well-specified consulting projects. They utilized their own special competence and experience from similar earlier projects. The competencies and processes in the field of architecture are well established, and certain standard ways of doing things have developed during the years. Information technology has changed the design tools, but the basis of the work has remained the same for a long time. This suggests that organizational routines may help to understand a lot of the everyday work in the architect's office.

As explained in section 2.4, routines help to automate recurring tasks and to store operational knowledge of the organization. The definition of routines was ambiguous with emphasis given to various parts of routines depending on the research context. In this thesis, routine work is seen as recurring, work related tasks, completed within certain time frame and which can relate to internal affairs or customer projects. In the questionnaire used, respondents themselves decided what was routine work related communication in their work. The questionnaire is in Appendix A.

Structural holes and brokerage between clusters is important for the emergence of new ideas. People located at the connections of two groups are exposed to new ways of thinking and behaving because opinions and behavior tend to be more homogeneous within than between groups. This is in line with the common wisdom that having dissimilar contacts to oneself gives raise to creativity and better ideas. New ideas arise when someone combines bits of knowledge across groups (Burt 2004).

A generic finding in sociology is that information flows better within groups than between groups (Burt 2001). Thus, it can be expected to see that densities within organizational units are higher than between the units.

In general, actors seek both, being embedded into dense groups and getting effectiveness from weak-ties. Task contingency view predicts that in tasks where the actor has the expertise required to understand the problem and its solution, actors don't need highly available resources for support. In fact, being highly connected is time consuming and starts to slow down the work (Hansen 2001).

These theoretical ideas suggest that the routine network should be denser and more clustered than ideas network, but not overly clustered to slow down the work. People are likely to communicate with those near them on routine work related matters.

It is reasonable to believe that project participants are involved into some amount of routine related information sharing in order to get the job done as efficiently as possible. The subgroup structure of project networks arises from the project data, not from social processes. The subgroup structure of the routine network probably reminds the projects network because of the project based work. Combining these ideas lead to the hypothesis that the project network is similar to the routine network, and they both are different to the idea network.

HYPOTHESIS 1: The project co-occurrence network derived from the project work hour data is structurally more similar to the network of discussion of routine work than to the network of discussion of ideas.

Dense interconnectedness causes knowledge to be redundant and new ideas are less likely to be discovered. Respondents are probably unlikely to state that they get the novel ideas from the persons who are working in the same group to themselves. Still, ideas are typically brainstormed and evaluated in groups for implementation.

The idea network can be expected to be sparse; it should contain structural holes and actors who reach for ideas outside their local clusters. Ideas are reached from outside using direct alters. This suggests that ideas network has ego-centric network structure (Smedlund 2010). As with communication networks, the idea network requires connectivity so that the ideas can spread from different parts of the organization.

HYPOTHESIS 2: Discussion of ideas forms structurally different network to the network of discussion of routine work and to the project co-occurrence network.

4 Research methods and data

This section starts by introducing the basic methodological concepts of social network analysis. Many key ideas of social network analysis are based on graph theory. Statistical methods require modifications because the independence assumption of observations does not apply to network data. In addition, the statistics used to describe social networks differ from the descriptive statistics commonly used in data analysis. There are also individual level measures that are unique to social network analysis. Dimension reduction methods such as multi-dimensional scaling and correspondence analysis are described toward the end of this chapter. The last two sections introduce the data used.

4.1 Social Network Analysis

Social network analysis offers precise definitions of concepts and a framework for testing theories about social relations. It formalizes statements and operationalizes network theories so that it is possible to measure properties of the network structure, to see trails of the theoretical concepts in the empirical data and to test hypotheses (Wasserman & Faust 1994; Brass et al. 2004). Social Network analysis does not just focus on expressing the existing sociological concepts formally, but many new theoretical concepts have emerged from the framework it offers for the study of social phenomena. (Borgatti et al. 2013)

Social network analysis is often used in studies to provide descriptive or explanatory insights to networks. Some studies aim to understand observations as a result of a causal process from initial conditions and theorized processes. The role of network analysis in these kinds of studies is to generate either the independent variables or dependent variables.

Network research designs are typically studies of whole networks or studies of personal networks. Studies of personal networks focus on the ego or the index node and its ties to other nodes, called alters. Studies focusing on ego networks typically aim to understand something about ego's social environment. Whole network studies on the other hand focus on the set of ties among all the nodes in the network. Ties can be social relations between nodes, interactions, transactions or flows of ideas, goods, information or infections. Ties can be constructed from co-occurrences of nodes or their attributes, such as the occurrence of same age or gender. Full network studies allow using wider set of tools and concepts of network analysis than ego-network studies. For example, positional concepts such as betweenness centrality assume that the entire network is available. (Borgatti et al. 2013, pp.28–29)

4.2 Gathering data

The most common method of gathering network data when actors are people is by questionnaire. In a standard sociometric data design, respondents are presented a roster of names of other people in the actor set and asked to rate their connections with each person. If actor set is not known beforehand, a free recall method can be used. Actors can pick names from the roster freely or if fixed design is used, they are told how many other actors to nominate on a questionnaire. Questionnaire usually has questions requesting respondent to rate or rank order the actors for each measured relation. This produces weighted network relations. These kinds of surveys utilizing questionnaires with questions of each actor's network ties are often called sociometric surveys in the social network literature. Other common ways to collect network data include interviews, observation and using archival records. (Wasserman & Faust 1994, pp.43–45)

From the social network perspective, individual behavior is seen to be contingent on the actor's social relationships and the network configurations in the social structure actor is embedded in. Network surveys differ from traditional survey methods because usually the goal is not just to get a representative sample of the population, but to gather as a complete picture as possible of all the relations between individuals. Getting accurate data of the whole network requires that researchers know where the boundaries of the network are. In addition, social network analysis requires care in specifying which social relationships should be studied. These concerns are often referred to as boundary setting problem in the network analysis literature. (Laumann et al. 1983)

There are two generally recognized strategies for approaching the boundary setting problem in network studies. Realist strategy of setting network boundaries assumes that actors themselves know whom they socialize with. In this case the boundaries can be taken to be defined by whom the actors include in their network and whom are left outside. This assumes that the network members have a collective awareness of all or most of the network members. Smaller groups are recognized by their members and in principle they have boundaries. In the nominalist approach researcher consciously sets the boundaries to serve the analytical purposes. (Wasserman & Faust 1994)

Network datasets contain data of actor level relations, sometimes also the strength or frequency of the relations and often the background attributes such as position, gender and age. The variables representing different relationships are referred to as structural variables (Wasserman & Faust 1994, p.3) and actor level attributes are often called compositional variables (Wasserman & Faust 1994, p.22). Relations that are collected can be directional pointing from the source actor to the target actor or unidirectional, just existing between actors. Dichotomous relations are either existing or absent and weighted relations include numeric measure of the strength, intensity of the frequency of interaction.

4.3 Graph representation of social network

Graph theory has served as the foundation for the most of the methods in social network analysis (Wasserman & Faust 1994, p.92). Ideas such as networks, groups, cliques and connections are often intuitively understood, but imprecise descriptions of the terms prevent from finding evidence of them in data or analyze them formally. Graph theory is the foundation to formalize these concepts. Social network analysis is not a part of graph theory, but many of the concepts and much of the vocabulary has been borrowed from this field. (Prell 2012, p.8)

Graph theory provides useful and simplified representations for social networks. A *Graph* is a mathematical object consisting of two sets of information: a set of nodes (or sometimes called vertices) and set of lines (or edges). Notation of graphs is based on sets and graphs are formally shorthand as a pair of set of vertices V and edges E or $G = (V, E)$. Self-ties or self-loops are commonly ignored. If there are undirected edges and no self-loops, then the graph is called a *simple graph*. *Nodes* in the graph represent actors, and *lines* represent social ties. Ties occur between actors in some relation, such as friendship, advice, communication or business transactions. Nodes are *adjacent*, if there is a line connecting them and the line is said to be *incident* with a node if that node is located in the either end of the line.

Graphs can be visualized by plotting the nodes as points, and lines incident to these nodes as lines between the points. Location of the points or length of the lines does not matter. What matters is the presence or absence of a line between the points.

In addition to a visual plot, graphs can also be described by matrices. *Adjacency matrix* or *sociomatrix* is a $g \times g$ square matrix for a graph with g nodes. Each element indicates whether a line is present or not. If a line is present between nodes i and j, the element x_{ij} in sociomatrix has a value of one. If the line is not present, it gets the value of zero. By convention, adjacency matrix is interpreted so that the node in a row sends something to the node in a column. In simple graphs the diagonal values are either set to zero or set as missing. The adjacency matrix is symmetric if a tie is present regardless is it from x_i to x_j or x_j to x_i.

Relations in social networks often have direction and weights. The directions can be represented with *directed graphs* or *digraphs*. Lines are often called *arcs* in digraphs. The adjacency matrix of a digraph is not symmetric unless the relations that arcs represent are reciprocated. In visualizations, the direction of the arc is indicated by an arrow and a two-way arrow indicates reciprocated arcs. Weighted graphs or valued graphs have a number assigned to each line. The numbers can represent the frequency of communication, money value, rankings, strength or intensity of a tie. A graph consists then of three sets: nodes, edges, and weights. Adjacency matrix of a weighted graph has weights instead of ones and zeros in each element. In a plot representation of a weighted graph, zero-weighted lines are typically not drawn. Many network properties are not as well developed for valued networks as for graphs and directed graphs. Same applies for many social network analysis methods. It is common to reduce valued network to simple graphs or digraphs for analysis purposes. (Wasserman & Faust 1994, p.140)

A Subgraph of a graph *G* is a set of nodes and lines that resides inside *G*. Sub-graphs are relevant to finding components and cliques from the network. Subgraphs can be generated by choosing nodes and then letting those lines to be part of the subgraph that are between selected nodes. Alternatively, one can pick a set of lines, and let the nodes that are incident to these lines to be part of the subgraph. Node-generated networks are commonly used to study groups in social networks. (Wasserman & Faust 1994, p.98).

A *walk* is a sequence starting from any node and ending to any node and is formed by following the lines between nodes. The length of the walk is the number of lines in the sequence. A *trail* is a walk with a limitation that each line occurs in the walk only once; a *path* has a further limitation that each node in the walk occurs only once. Two nodes are reachable if there exists a path between them (Wasserman & Faust 1994, pp.105–107). As a practical example of the difference of the walk and paths, gossips and jokes are not usually told to same person twice and follow paths in a social network, viruses and infections can visit same person twice and spread along walks. If the graph is directed, directions of the arcs can be taken into account in walks.

A graph is *connected* when every node in it is connected by a path. If a graph is not connected, it is *disconnected,* and it is partitioned into separate components where there are no path between the components. *Component* is a maximal connected subgraph. There are no edges connecting other nodes and the nodes in a component. A Component may consist of just a single node which in this case is often referred as an *isolate*.

There can be several paths connecting any two nodes in the graph. The shortest of the paths between two nodes is called a *geodesic*. Geodesic distance refers to the number of lines along the geodesic. Usually, this is just called a *distance* between two nodes. If there are no path existing between two nodes, then their distance is infinite or undefined. The largest geodesic distance in the whole graph is called the *diameter* of the graph. If the graph is disconnected, the diameter is infinite or undefined. (Wasserman & Faust 1994, pp.110–112) Often in applications diameter is the largest geodesic distance of the largest component.

Reachability along only one path makes some nodes and lines critical for the connectivity of the graph. A node is called a *cutpoint* if removing the node causes the graph to have more components. In a social network, a cutpoint actor is critical for information flow from one part of the network to another. This kind of actor is also powerful in the sense that he is in a gatekeeper position to regulate the flow of information. If several nodes must be removed before the number of components increases, then these nodes are said to form a *cutset*. Similarly, a critical line is called a *bridge* if it connects two otherwise separate components. A bridge represents a critical connection in a social network that the actors have to maintain.

4.4 Comparing networks

Social network analysis offers different sets of methods depending on the network level. A lot of social network studies have focused on structural positions of actors. Network variables are typically explanatory variables explaining various outcomes. Examples of these types of studies are the social capital studies, where actor's performance is explained by the properties of his network (Burt 2002).

If two different kinds of relations are compared within the same set of actors then there are a large amount of methods available. There are plenty of statistical methods developed from whole network level descriptive statistics to comparative methods such as QAP-correlations (Wasserman & Faust 1994). Correlation can be used to see if the ties of two networks tend to occur together. Positional differences of all actors can be compared with dimension reduction such as MDS or correspondence analysis. Centrality measures can reveal differences how the same actors are embedded differently in the networks being compared.

Recently, methods have been developed that can be used to compare networks on structural properties regardless of the set of actors, the type of ties or the size of the networks.

Exponential random graph models (ERGM), also known as p* models are a class of statistical models for social networks that can be estimable from data and are empirically grounded (Robins et al. 2007). Networks can be considered to arise from local configurations such as dyads, triads, stars and cycles and social processes such as homophily, transitivity and preferential attachment. By including these parameters, a model can be estimated based on social processes (Lusher et al. 2012). Exponential random graphs can be fit to various seemingly different networks. Faust and Skvoretz (2002) demonstrated the utility of the ERGM by comparing very different kinds of network from co-sponsorship among senators to the network of social licking among cows and grooming among monkeys (Faust & Skvoretz 2002).

The relations studied in this thesis are different relations in the same set of actors. There are multiple network analysis methods available for graph level, actor level, subgroup level and dyadic level. The underlying relations being studied are symmetric.

4.5 Centrality measures

One of the most common questions in social network analysis is who are the most important people in the network. In order to answer that question, several actor centrality measures have been developed. It turns out that the importance of an actor depends on the situation. The uniting idea behind different approaches is that important actors are located in strategic locations in the network. Actors must either have many connections, or connections to powerful actors, or the actors themselves must be located either in between other actors, or near as many actors as possible. These different ways of seeing actor importance lead to different kinds of centrality measures.

4.5.1 Actor degree centrality

The degree of a node in a graph is the number of lines that are incident with it (Wasserman & Faust 1994, p.100). If the graph is directed, it is common to make a difference between the degree resulting from arcs pointing outwards from the node and arc pointing inwards to the node. *Indegree* is the number of nodes that are adjacent to the node; *outdegree* is the number of nodes that are adjacent from the node (Wasserman & Faust 1994, p.126).

Degree centrality is the sum of adjacent nodes, or in the adjacency matrix of a simple graph, the row sum.

$$C_D(n_i) = \sum_j x_{ij} = \sum_j x_{ji}$$

If the group or graph has g nodes, the maximum degree an actor can get is $g-1$. The degree centrality measure is often standardized by dividing by the maximum degree $g-1$.

In a simple graph the directions of the lines do not matter: there either are incident lines or there are not. In social networks, this is same as the number of actors one has ties with. Degree centrality measure pays attention only to direct or adjacent choices.

Degree of a node commonly draws a lot of attention in social network analysis. High degree in the graph shows that the actor has a large number of ties in the social network, and this is aligned with the idea that central actors in a network must have a lot of social activity. Outdegree measures expansiveness and indegree measures popularity (Wasserman & Faust 1994, p.126). In the friendship networks, a person with a high indegree receives many friendship nominations from the other actors. Sometimes network questionnaires are designed so that the outdegree is fixed. For example, respondents could be asked to name their three best friends. (Wasserman & Faust 1994)

4.5.2 Actor closeness centrality

Centrality can also be considered based on how close an actor is to other actors in the network. In this case, the actor is central if he can quickly interact with others. Being close to other actors can make these actors good at communicating in the network. An actor with high closeness centrality can easily access information from other parts of the network. In that way, closeness centrality gives a good description of centrality that is consistent what centrality intuitively means. Centrality is also global measure that takes into account the whole network and not just the neighborhood (Perry-Smith & Shalley 2003).

An actor is central from the closeness centrality point of view if the geodesics, or shortest paths, linking it to the other actors are as short as possible. In other words, the longer the distances to other actors, the less central the actor is. Sabidussi (1966) defined index of actor closeness as an inverse of the sum of the distances to the other actors.

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1}$$

Subscript C refers to closeness centrality. Closeness centrality can vary from zero when an actor is not reachable to $(g-1)^{-1}$ if the actor is adjacent to all actors. (Wasserman & Faust 1994, p.188; Freeman 1979)

4.5.3 Betweenness centrality

If an actor is located between two nonadjacent actors in a network, he can potentially have control over how the other two actors interact. If a node is central in the network, it should also be central compared to all other nodes in the network. With this view, the focus is not in how many people one knows, but where one is located in the network. (Prell 2012, p.104)

A node in the graph is central if it is located between other nodes on their geodesics. Large betweenness centrality mean that the node lies between many actors on their geodesics. If all geodesics are equally probable paths for spreading of information, probability of choosing any one of the possible paths between two nodes is $1/g_{jk}$ where g_{jk} is the number of paths. If $g_{jk}(n_i)$ is the number of geodesics containing actor i, probability of passing through node i is then $\frac{g_{jk}(n_i)}{g_{jk}}$. Freeman betweenness centrality is calculated as the sum of probabilities over all pairs of actros not including actor i.

$$C_B(n_i) = \sum_{j<k} g_{jk}\,(n_i)/g_{jk}$$

Betweenness centrality can be calculated even when the graph is not connected. Betweenness can vary from the minimum of zero to maximum of all pairs of actors except the node i itself, or $(g-1)(g-2)/2$. (Wasserman & Faust 1994, p.189; Freeman 1979).

4.5.4 Eigenvector centrality

Eigenvector centrality can be seen as a weighted modification of centrality degree. It extends the degree measure by taking into account that ties to actors with many connections are more valuable than ties to actors with fewer connections. With this view, what degree centrality does is the same as eigenvector centrality but it gives equal weight to all adjacent nodes. The derivation and the formula for eigenvector centrality are in Appendix B.

Eigenvector centrality works well for simple graphs as its adjacency matrix is symmetric. For asymmetric matrices, formulas must be modified to take into account the two set of eigenvalues. Other measures for node importance are Katz centrality and Google PageRank. Both of them work similarly to eigenvector centrality with some adjustments related to directional networks. (Newman 2010)

The four different centrality measures discussed here can be summarized by their different points of view they take to centrality. Degree centrality can be seen as measuring activity of an actor, eigenvector centrality measures the extent an actor is surrounded by highly connected actors, betweenness measures potential control for information flow and closeness centrality indicates actor's independence of information channels. (Prell 2012, p.108)

4.6 Cohesive sub-groups

Social networks are typically not evenly dense, but people tend to cluster and form denser areas in networks. There are various names for the dense clusters, including cliques, clans and social circles or more formally, cohesive subgroups. Cohesive subgroups are not vulnerable to removal of ties as there are many paths connecting nodes. (Wasserman & Faust 1994)

For the analysis purposes, it is desirable to have definitions for the subgroups which make it possible to assign actors unambiguously to different subsets. Groups and communities in social networks do not usually have clear boundaries, but they overlap. The group boundaries drawn in social network analysis often result from the methods used, such as hierarchical clustering (Kadushin 2011, p.90)

A clique is a maximal complete subgraph where it is not possible to add other nodes while keeping all the nodes in the subgraph fully connected. There are various extensions for cliques that relax some of the conditions, such as k-cores. K-cores are groups of actors connected to some number (k) of other members. (Hanneman & Riddle 2005; Scott 2013)

Core/periphery structure is one of the simplest form of network segmentations, and very familiar from everyday experience. There are people who are "inside" and then there are the outsiders. The concepts of core and periphery have been used quite informally without giving formal definitions what this means (Borgatti & Everett 2000). The idea is to classify the actors into two groups: to a relatively dense core, often connected by strong ties, and to a sparser periphery, connected mostly with weak ties. In addition, networks have often been assumed to have only one core surrounded by peripheral nodes (Borgatti & Everett 2000).

There can be many different kinds of cores, depending on what kind of interaction there are between the core and the periphery. The core might have interaction mainly between the actors in the core, or there can be diffusion from the core through unreciprocated ties to the periphery. In Addition, there can be interaction inside the periphery or there can be situations where core does not have to interact with anything. In this case the only interaction in the network is in the periphery. (Kadushin 2011, p.100)

4.7 Whole network descriptive statistics

Various network measures are used to describe networks, in the same way as descriptive statistics are used in conventional statistical analysis to give an overview of the data.

Density is perhaps the simplest and most used cohesion measure in network analysis (Wasserman & Faust 1994, p.181). In a network with g nodes, there are $g(g-1)/2$ possible unordered pairs of nodes. Density is calculated as the number of ties present as a proportion of all the possible ties, or $\Delta = L/[g(g-1)/2]$ where L is number of edges and g is number of nodes in the graph. Density can be interpreted as how connected the network is compared to a fully connected network. People have only limited time to keep up with growing number of connections and density level tends to get lower in a social network as network size increases (Borgatti et al. 2013, p.151).

Average degree gives similar information to network density but in a different form. Average degree tells how many connections actors have in average in the network. It is calculated from Freeman degree centrality measure as a simple arithmetic average over g number of nodes or $\overline{C}_d = \sum C_d(n_i)/g$. If the measures is standardized to range from zero to one by dividing by $g-1$ then the measure is exactly same as density discussed above (Wasserman & Faust 1994, p.102).

Clustering coefficient measures the extent to which the network has areas of higher densities and lower densities. One recurring idea related to social networks is information spreading out from a single node to the network. If the network is well connected, the whole network will eventually get the spreading information. The question is really how many steps it takes. The concept of neighborhood captures the idea of these steps. The neighborhood $\Gamma(v)$ of a node v is a subgraph which consists of nodes directly connected to v. Clustering coefficient γ_v is calculated as the ratio of the number of edges in the neighborhood of the node v and the total number of possible edges in the neighborhood, or $\gamma_v = E(\Gamma_v)/\binom{k_v}{2}$. Clustering coefficient of the whole graph is the average valued of node clustering coefficient and gives the average density of all open neighborhoods in the network (Watts 2004, p.31). Clustering coefficient can also be calculated as a proportion of transitive triads (in an undirected network, groups of three actors with three connections) to all connected triads.

Social networks tend to have clumps of actors clustering together and small path lengths (Borgatti et al. 2013, pp.156, 260).

Freeman centralization measure measures how central the most central node in the network is compared to all other nodes and expresses this as a ratio to the theoretical maximum in a binary network. it is calculated as the sum of differences between each node's degree centrality and highest observed degree centrality, divided by theoretical maximum possible coming from a node connected to all other nodes or in algebraic form as in equation below. (Freeman 1979)

$$C_x = \frac{\sum_{i=1}^{N} C_x(p_*) - C_x(p_i)}{max \sum_{i=1}^{N} C_x(p_*) - C_x(p_i)}$$

4.8 Statistical methods

Graphical representation of networks can give a lot of information about the underlying social structure. However, when the network contains tens of nodes, the graph visualizations become difficult to interpret. Statistical methods help to reduce subjectivity from the interpretation and describe network properties when it is not possible anymore simply by visualizing.

4.8.1 Hypothesis testing with correlated observations

Dyadic level correlations between matrices can be performed in an ordinary way by reshaping the matrix into a long vector and calculating normal Pearson correlation co-efficient. Standard statistical tests assume that observations are statistically independent and drawn from a population of a particular distribution. In social network data, observations are not independent. For example, if a person reports communicating daily with another person, one can expect that the latter person also reports communicating with the first person. As network observations are not independent, and the observation space is the whole population with unknown distribution of ties, standard statistical significance tests cannot be used.

Quadratic Assignment Procedure (QAP) is a permutation-based and nonparametric test that preserves the autocorrelations in the matrix. It works by shuffling both, the rows and columns of the adjacency matrix simultaneously and correlating the observed matrix against the resulting permutated matrices. For example, in a 5 actors networks there are 120 (=

5!) possible ways of the five actors to form a network. The number of possible permutations grows quickly with the network size. Usually, the test is run against a sample of 5 000 to 50 000 permutations. The p-value of the statistics is the proportion of how many times the observed correlations resulted with different permutations of the matrix. (Krackhardt 1987; Borgatti et al. 2013, p.126)

4.8.2 Multidimensional Scaling (MDS)

Multidimensional scaling (MDS) is a set of techniques used mostly for data visualization. MDS attempts to arrange observations to a space with chosen number of dimensions so that distances between points in the space correspond to original proximity data. Related variables in input proximity data are mapped so that they are near each other in output coordinates. Proximity data can be correlations, measures of similarity, co-occurrences and so on.

Classical MDS, also known as Torgerson-scaling, originates from 1950s. It is done directly by matrix calculation and is based on Eigen-decomposition of the proximity matrix. In classical MDS, proximities should be presented as distances. Distance from an object to itself is zero, distance is symmetric between objects ($d_{ij} = d_{ji}$) and distances satisfy the triangle inequality $d_{ik} \leq d_{ij} + d_{jk}$. In the field of psychology, a family of distance functions are used. A generalized Minkowski distance has the following form:

$$d_{i,j}(X) = \left(\sum\nolimits_{a=1}^{m} \left|x_{ia} - x_{ja}\right|\right)^{1/p}, p \geq 1$$

If $p = 1$, distances are called city block distances, $p = 2$ gives familiar Euclidean distances, and as $p \to \infty$, formula results in what is called a dominance metric.

Illustration below shows classical MDS of inter-town distances that happen to be already Euclidean distances. Arriving at a set of coordinates required a series of matrix multiplications and eigenstructure computation. Resulting map is very similar to the map of how the towns are located in the real world. In this case, there are some visible distortions in the map. The input data consists of distances measured from the road network instead of great circle distance. In addition, distances are measured on a curved surface which MDS projects to flat 2D-presentation.

Data arising in most applications are not distance data, and more flexible iterative algorithms are used. Iterative algorithms allow rank orders for the proximities of the observations. Iterations begin with some starting configuration and move the data points to some directions. The new locations are compared to original distances using a stress function. Data points are then moved to directions that reduce stress measure. This process is repeated until stress value between iteration reduces to a less than some set threshold limit. Iterative algorithms cannot guarantee to find the global optimum and might get stuck to local optimums. In this case different starting configurations can be used to find other optima (Borg et al. 2013, p.6). When rank orders are used instead of distances, MDS is called ordinal MDS or non-metric MDS.

Amount of distortion, or goodness of fit of MDS solution is measured using Shepard-diagram or stress-function. Shepard diagram plots distances in the resulting map against proximities in the original data. There are many variants of stress functions, depending on the algorithms and software package used. The basic measure is called Stress-I or "Kruskal's Stress" and it is an aggregate measure of error terms of a least-squares regression line fit in the Shepard diagram and distances in solution. MDS algorithms try to minimize the total stress measure. Stress-I function is shown in equation X. The last division is done to normalize the measures between zero and one.

$$\text{Stress-I} := \sqrt{\sum_{i<j} \left(\widehat{d_{ij}} - d_{ij}(X)\right)^2 \Big/ \sum_{i<j} d_{ij}^2(X)}$$

An MDS solution with the perfect fit has a stress value of zero. A perfect fit doesn't have distortions in the resulting map. It is not possible to see from the stress value if the stress is caused by some points having a large error or many data points having small errors. Similarly to stress-I function, most different stress functions minimize squared residuals. This means that minimizing stress function gives more weight on minimizing largest data point distances. Large distances thus tend to be relatively more accurate than shorter distances. (Borgatti et al. 2013, p.91)

How much stress is allowed depends on the situation. Stress value does not tell if the background theory is compatible with the data, it just tells the total size of the error terms. The amount of the acceptable stress must be less than what would arise from random data. Simulation studies show that MDS of random data to 2 dimensions gives stress values order of 0.2 in a simulation with 12 objects and 0.3 in a simulation of 48 objects (Spence & Ogilvie 1973). By convention, stress values less than 0.12 are acceptable for non-metric scaling (Borgatti et al. 2013, p.91).

Stress values decrease with added dimensions as the new dimensions allow algorithms more freedom to position data points. Typically two dimensions are chosen for a flat presentation, but the number of dimensions can be decided with the help of scree-plot which plots stress value against each number of dimensions. An elbow in the scree-plot indicates the point where additional dimensions represent random components.

MDS maps are usually searched for dimensions or axes, regions and clusters of observations. Orientation of the axes is arbitrary. Clusters of data points that are near each other share some similarities. Data points can also be scattered on a continuum on an axis. When MDS configurations are compared, lot of meaningless information must be discarded. These include any similarity transformations such as rotation, translation, reflection and global enlargement or shrinkage.

	Helsinki	Tampere	Turku	Kuopio	Oulu	x	y
Helsinki	0	176	165	382	607	-211	-74
Tampere	176	0	157	293	476	-92	33
Turku	165	157	0	449	620	-240	84
Kuopio	382	293	449	0	284	164	-111
Oulu	607	476	620	284	0	379	68

1. Distances: $\boldsymbol{\Delta}$
2. Squared distances: $\boldsymbol{\Delta}^2$
3. Centering matrix: $\boldsymbol{Z} = \boldsymbol{E} - n^{-1}\boldsymbol{1}\boldsymbol{1}'$
4. Centering: $\boldsymbol{B} = \frac{1}{2}\boldsymbol{Z}\boldsymbol{\Delta}^2\boldsymbol{Z}$
5. Eigen structure: $\boldsymbol{B}_{\boldsymbol{\Delta}} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$
6. Coordinates: $\boldsymbol{X} = \boldsymbol{Q}_{+}\boldsymbol{\Lambda}_{+}$

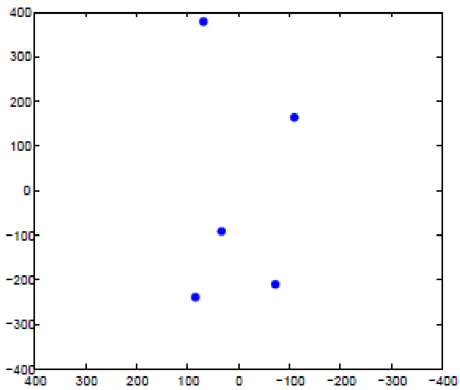


Figure 1 Illustration of MDS with distances between towns. Distances in the table form matrix Δ. After going through the computations listed from 1 to 6, two vectors corresponding to largest eigenvalues give the coordinates of the towns.

4.8.3 Correspondence analysis

Correspondence analysis (CA) is a descriptive and exploratory data analysis technique for studying correlations among two or more sets of entities. It is used for two-way frequency tables containing some measure of correspondence between the rows and columns. There are different variants of correspondence analysis but the goal is the same, to represent the data table as points in two-dimensional space (Wasserman & Faust 1994, pp.334–343; Knoke & Yang 2008, pp.113–117)

Suppose that the measures of actor centralities have been presented in a table where columns show four centrality measures and the rows list the actors. The columns can be thought to form four-dimensional space where the actors represent different points. One could calculate the distances between the points in four dimensions to measure similarity. Instead of representing the data in four-dimensional space, singular value decomposition (SVD) can be used to reduce dimensions to two dimensions. Before running a SVD, the data matrix is normalized by dividing the elements by the square root of the product of the corresponding of rows and column sums. These steps can be expressed in matrix notation by $\mathrm{R}^{½}\mathrm{F}\mathrm{C}^{½} = \mathrm{UDV}$. Where R and C contain the row and column sums, U and V give the row and column vectors and D is diagonal with singular values (Borgatti et al. 2013, p.92; Knoke & Yang 2008, p.113)

Singular value decomposition of the normalized data matrix produces three sets of scores: row and column scores and a set of singular values. The singular values show the importance of each of the dimensions. Scores for rows and columns are plotted as row and column points. The distances between the row points in the resulting plot measures the similarity of the row profiles in the original data matrix. Similarly distances between column points measure the similarity of column profiles. Row and column points are not directly comparable, but row points can be seen to be located in the neighborhood of the column points where the row profile is prominent. (Knoke & Yang 2008, p.114)

Figure 2 shows an example of data reduction using Doctorates dataset in UCINET. Points whose rows have similar profiles tend to be near each other, for example, Mathematics and Engineering. Points of similar years are near each other as well. (Borgatti et al. 2013, pp.92–95)

	1960	1965	1970	1971	1972	1973	1974	1975
Engineering	794	2073	3432	3495	3475	3338	3144	2959
Mathematics	291	685	1222	1236	1281	1222	1196	1149
Physics	530	1046	1655	1740	1635	1590	134	1293
Chemistry	1078	1444	2234	2204	2011	1849	1792	1762
Earth Sciences	253	375	511	550	580	577	570	556
Biology	1245	1963	3360	3633	3580	3636	3473	3498
Agriculture	414	576	803	900	855	853	830	904
Psychology	772	954	1888	2116	2262	2444	2587	2749
Sociology	162	239	504	583	638	599	645	680
Economics	341	538	826	791	863	907	833	867
Anthropology	69	82	217	240	260	324	381	385
Others	314	502	1079	1392	1500	1609	1531	1550

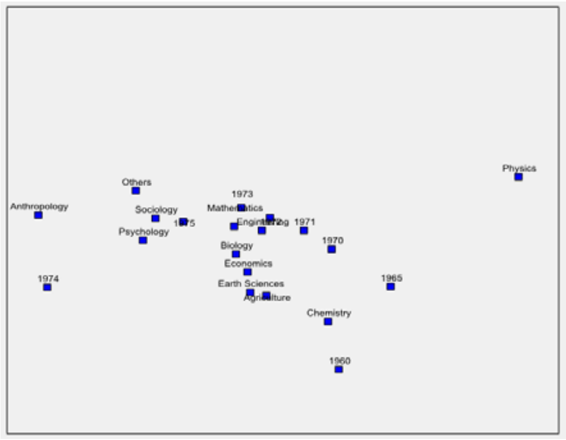


Figure 2. Illustration of correspondence analysis using Doctorate-dataset in UCINET. Correspondence analysis gives a visual representation of the otherwise hard to read data table.

4.8.4 Hierarchical clustering

Hierarchical clustering (Johnson 1967) is used to build a hierarchy of clusters by assigning observations into mutually exclusive groups. Agglomerative algorithms progress bottom up by starting from the observations in their own clusters and by merging the clusters as the algorithm moves up the hierarchy. Divisive algorithms progress top down by starting with all the observations in one cluster and by splitting the clusters recursively until all the observations form their own clusters.

In social network analysis, one application of hierarchical clustering is to identify groups or actors who participate in many groups. A proximity matrix is produced by MDS or other means. An agglomerative algorithm progresses in the following manner: First, all observations are in their own cluster of size one. Next, the pair having closest distance between them is merged together. Distances are then calculated again between this merged group and the other observations by taking into account the nearest distance of the merged group. The nearest observations within the group are merged, and the process is continued until there is one large group consisting of all the actors. The results are often shown in a dendrogram, a tree diagram showing the partitioning of the data at different steps. (Borgatti et al. 2013)

4.9 Two-mode networks

People often form social connection to other people by participating in the same events. It is possible to build a network representation of event members by connecting a pair of actors if they participated in the same event. A network consisting of two sets of nodes, people and the events is called two-mode network or affiliation network. Similarly, as people can be linked by the events they participate in, events can be linked in terms of people. This is often referred to as the duality of persons and groups (Breiger 1974). Strengths of ties in two-mode networks are often defined in terms of the frequency of participation.

Two-mode networks can be studied by using bipartite graphs or by converting them to one-node networks. Bipartite graphs are graphs that can be partitioned into two sets of nodes so that all the ties in those sets are formed only within set (Wasserman & Faust 1994, p.120). Although using bipartite networks to analyze two-mode networks retains much of the network properties, there are a limited set of social analysis methods extending to two-mode networks. For that reason, two-mode networks are often projected into one-mode networks. (Hanneman & Riddle 2005)

To avoid loss of information through one-mode conversion, sometimes two different networks are formed: one for actors, one for the events. Loss of information is probably not a serious issue if the focus is only on one of the modes.

Two-mode networks can be converted to one-mode networks by connecting a pair of the nodes in one node-set if they both are connected to the same node in the other set. Figure 3 illustrates the projection of a two-mode bipartite graph onto one-mode graph. Projecting two-node network to one-node network can be done by multiplying the co-occurrence matrix by its transpose.

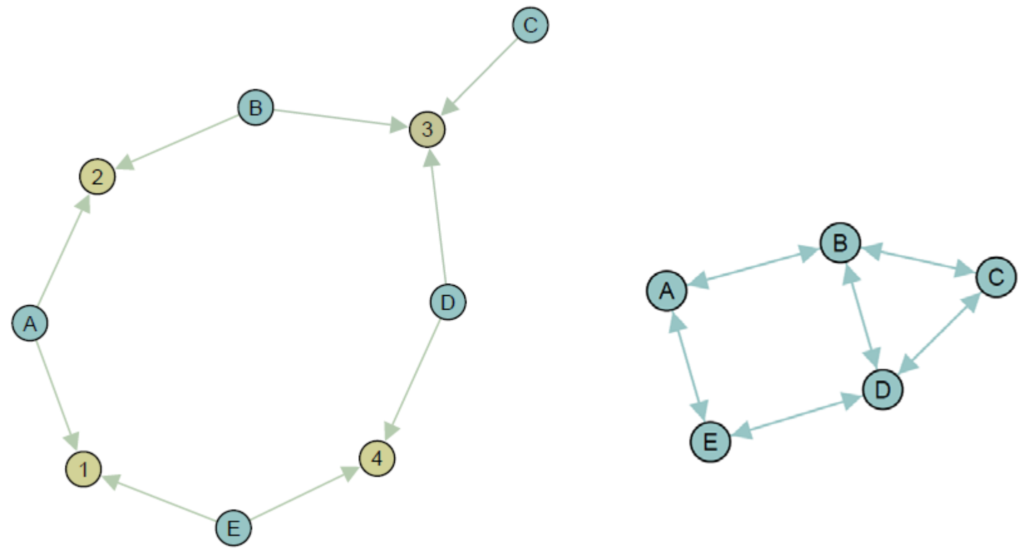


Figure 3 Projecting bi-partite graph onto one-mode graph. Nodes with alphabets represent actors, nodes with numbers represent projects. A tie is inferred between actors if they participate in the same project. For example, a tie is formed between actors B, C and D who all participate to project 3.

4.10 Datasets used in the study

The datasets used in this study were created by Anssi Smedlund in 2007. The data comes from a study conducted in an architect's office in Northern Europe (Smedlund & Choi 2009). Structural and compositional variables were gathered using sociometric survey. Employees of the architect's office were mostly professionals, and certain job positions required Master's level or higher degree in architecture. The case company can be characterized as a professional service firm, and their work is highly knowledge intensive. The work in the case company was project based, and the employees worked on various architect projects either on client projects or on firm's internal development.

The data was collected for the study of intra-firm networks and consist of self-reported network ties of discussion of routine work related matters and discussion of ideas. In addition, a dataset consisting of the reported total work hours for each project was collected. Each employee in the case company kept close track of their working hours. Work reported for year 2007 consisted of billable work, internal development work and general office work marketing activities. (Smedlund & Choi 2009)

Employees in the network had five different roles. Most of the employees were professionals (N=49) working on drawings of buildings and public spaces. Managerial tasks were handled by project managers (N=17), senior project managers (N=13) and middle and top managers (N=9 and N=6). There were also employees in administrative roles (N=14).

This thesis uses the both datasets mentioned above: The self-reported network data of discussion of routine and non-routine communication and the data of reported project hours. Background variables were also included in the dataset, but here they are used only to help to clean and pre-process the data. The analyses in this thesis focus only on the structure of the network ties and network positions of the actors. Organizational unit is used as a background variable to check differences in network densities between units.

Self-reported Networks

The organization of the architect's office consisted of several different units. Some of the employees were external workforce and the company had offices in different countries.

The boundaries of the data collections were set to include the actors who were based in the same office. The questionnaire presented a roster, where everyone in the office was included. Respondents defined their communication network by selecting those persons from the list whom they recalled having communicated with during the past year.

After the respondents had defined their communication network, they were presented questions of whom they discussed on routine work and ideas with. The survey took around 30 min for the participants to complete. The data produced is multi-relational, valued and undirected sociometric data. 77 persons answered the questionnaire and named 109 persons into the social network. Survey questions are presented in Appendix A.

4.10.1 Project data

The project dataset used in this thesis consists of a list of all the projects the case organization worked on during year 2007 and the total work hours each employee reported to each project. Even though the project data is not network data, the project data can be used to construct network relations between each employee. Information on the employees and the projects they worked on can be presented as a two-mode network of actors and projects. Two-mode network can be converted to a one-node network by using the methods described in section 4.9. The conversion results in a third one-mode network of the same actors.

In this thesis, the networks of discussion on routine work and ideas and co-occurrences in projects are later referred to as the routine network, idea network and project network.

4.11 Data preparation

Network data has to be formatted to requirements of network analysis packages. Network analysis was done using UCINET 6 (Borgatti et al. 2002).

The routine network, the idea network and the project network are logically undirected networks. Even though the underlying social relations of discussing on routines or ideas are symmetric, it does not mean that respondents have exactly similar memory on whom they discussed with and how often. Different answers from respondents create asymmetry into the adjacency matrices. In this case it makes sense that the directions of the network ties are ignored. Methods such as multidimensional scaling explicitly require symmetric data. Some methods such as quadratic assignment procedure work better if the both relations being compared are either symmetrical or directed relations (Borgatti et al. 2013).

Many network analysis methods require binary data where the tie is coded either existing or absent without any indication of the strength of the tie. There are analytical tools existing for valued networks, but the methods coming from graph theory are well founded and more commonly used in social network analysis. The process of reducing a valued network to a binary one consisting of only ones and zeros is called dichotomization. Unfortunately, there seems to be no one agreed method on how to binarize valued network. In addition, dichotomization decision may have an impact on the network analysis results (Thomas & Blitzstein 2011b; Thomas & Blitzstein 2011a).

Dichotomization problem is easy to deal with if the valued network data is in an ordinal scale. In that case, the analysis can be limited to the strongest ties as usually the most frequent interactions are in the focus of the study. When the values are measured on a continuous scale, the dichotomization decision is more challenging as the data provides very little hints where the correct cut-off value should be.

4.11.1 Symmetrization

The raw data of the routine network and the idea network were asymmetrical. Respondents mentioned discussing their routine work and ideas with certain people in the organization, but were for some reason not mentioned by their counterparts. The proportion of reciprocated ties of all ties in the network was 0.31 for the routine network and 0.28 for the idea network data. The underlying relations that were probed with the questionnaire however were considered as symmetric relations.

Asymmetry could arise from reporting errors simply by forgetting some interactions. If either of the actors share information, then there exists an exchange tie. Two-way questionnaire design can be used to reduce some reporting errors. For example, in the case of routine network, two questions could be asked: “who do you give information related to routines to” and “who do you receive information related to routines from”. These questions probe the same underlying relation of information exchange, but from two different directions. Answers can then be combined into a single tie by transposing the matrix of either of the answers and taking an average of the two. (Wasserman & Faust 1994, p.157). Unfortunately, two-way questions make questionnaires longer to fill for respondents, and response rate could be lower because of respondent fatigue. For this reason, shorter question of the existence of a tie was chosen.

In order to keep the four-step response scale intact, the routine work network and the idea network were symmetrized by taking the larger self-reported value to indicate the strength of the information exchange of that pair.

4.11.2 One-mode projection of project data

Next, the network of participating to projects was constructed from the project working hour data. Only the subset consisting of the hours reported by the survey respondents was kept, and the rest of the data was discarded. Work hours data contained holidays as a separate project and these entries were also removed.

The construction of the project network rests on the idea that there is an underlying one-mode communication network under the two-mode network formed by project participants and projects. The extra step of converting the two-mode network explicitly to a one-mode network reveals the underlying communication network, which can then be compared to discussion networks of routine work and ideas. Actually, it is not possible to know if the projects participants really communicated with each other, potential for communication would be more descriptive of the characteristics of the network than communication network.

The project hour data is easy to convert to a weighted actor-event network (two-mode network) consisting of survey respondents and the project they reported having worked on. This was then converted to a one-mode network by forming a tie between a pair of actors if they reported to the same project. The resulting one-mode network is symmetric due to the matrix multiplication as explained in chapter 4.9.

Projection on one-mode network requires a slight modification from the direct matrix product. A direct matrix multiplication works well with binary network matrices. However, in the case of weighted adjacency matrix, the values of the resulting product matrix are hard to interpret. Instead, a minimum of the reported hours was chosen as the value of the product matrix. With this approach, each element in the product matrix indicates the maximum amount each pair of persons could have worked with each other. The resulting values of the one-mode network ties can be interpreted as the strength of the potential to communicate. Figure 4 illustrates the procedure.

The adjacency matrix of two-mode network:

	1	2	3
A	0	3	0
B	4	6	0
C	5	0	7

Projection by multiplication: $AA^T = \begin{pmatrix} 0 & 3 & 0 \\ 4 & 6 & 0 \\ 5 & 0 & 7 \end{pmatrix}\begin{pmatrix} 0 & 4 & 5 \\ 3 & 6 & 0 \\ 0 & 0 & 7 \end{pmatrix} = \begin{pmatrix} 9 & 18 & 0 \\ 18 & 52 & 20 \\ 0 & 20 & 74 \end{pmatrix}$

Projection by cross-minimums = $C_{ij} = \sum_k \min(A_{ik} A_{kj}) = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 10 & 4 \\ 0 & 4 & 12 \end{pmatrix}$

Figure 4 Projecting two-mode networks into one-mode networks by i) a direct multiplication and by ii) cross-minimums method. The advantage of the latter method is that the resulting matrix has an easy interpretation as potential of communication.

4.11.3 Dichotomization problem

Dichotomization refers to converting weighted networks into networks with just zeros and ones in their matrix presentation: one indicates that a tie is present between actors, zero indicates the absence of a tie. Dichotomization is done for analytical purposes because most of the analytical tools in social network analysis require binary data. In this study, dichotomized networks were created of the routine, idea and project networks.

The goal of using analytical methods is to learn about the underlying network, but moving from weighted networks to binary networks necessarily loses information. The dichotomization rule should in some way be chosen to minimize the loss of information. A concern with dichotomization is that the chosen threshold level might affect the results and could lead to wrong inference. For example, a network could have one giant component connected loosely with weak ties. Settings threshold level too high would zero out some of the weaker ties and break up the giant component into parts. Too high a cut-off value also tends to straighten ring-shaped networks and alter the network's topology. In both cases, the characteristics of the whole network would change radically. On the other hand, setting a cut-off value can also act similarly to a high pass filter in electronics, passing the high-frequency interaction in the network and reducing the noise by zeroing out the infrequent relations. Setting the cut-off value too low adds noise to the analysis and can prevent inferences of the data. (Thomas & Blitzstein 2011b)

Anderson, Butts and Carley (1999) show that the density of a graph affects graph level measures such as degree centralization, and these measures may have different interpretations depending on what density the dichotomization cut-off value caused. Unfortunately, literature doesn't provide any agreed methods or decision rules about how to choose the dichotomization threshold.

There are at least two commonly used methods to dichotomize a weighted network. One way is to use an uniform cut-off value and set all the values under the cut-off value to zero and all the values above this value to one. Bernhard, Killworth and Sailer (1970) used this approach in their informant accuracy studies. The objective of their study was to find out how observed and self-reported network ties correspond to each other. They set dichotomization threshold to minimize the difference between the recalled network and the observed network. In practice, this meant that the densities of the observed and recalled network were nearly equal.

Another approach is to choose each actor's *k*-strongest ties as the binary network and re-symmetrize the network if necessary. Thomas and Blitzstein (2011a) show that this method, in addition of being more complicated, turns out to be not any better than the first one, and performs even worse.

For this thesis, the first mentioned approach of setting uniform cut-off value over whole network was chosen because it is the most common way to dichotomize network data and is readily done in social network analysis packages without further programming.

4.11.4 Dichotomization of routine, idea and project networks

Routine network and idea network are easy to dichotomize as the data has been collected in an ordinal scale. Most of the analysis in this thesis are done using the daily frequency of communication. Ties with daily interaction represent the most frequent communication and probably the strongest ties as well.

Project data consist of continuous ratio data of the total hours each person reported for projects. The four biggest projects in the data were used to lump together various kinds of work that could be categorized as office work, development, marketing or risk management. Everyone in the organization reported some office hours, but certain types of employees such as support assistants reported only office hours. Fixed hours, limited hours and unlimited hours were top level categories of a large number of different client projects. The work hours by different categories are shown in Figure 5.

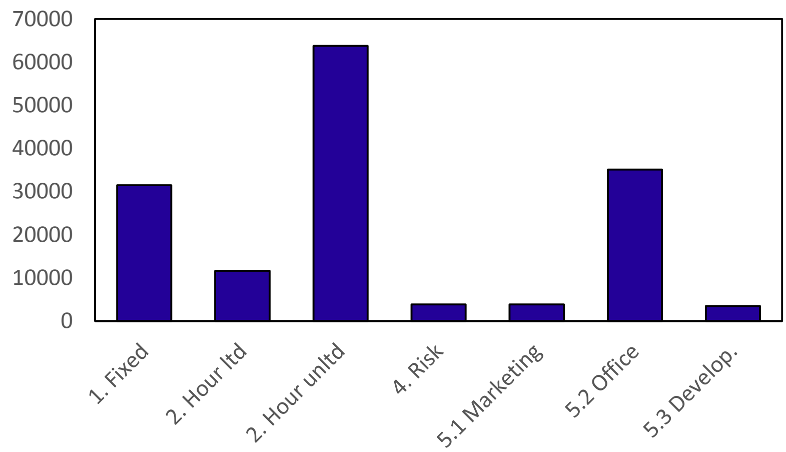


Figure 5.The project hours were categorized by the architect's office into different billing classes. Around 70 % of the work was done in projects either by fixed hours, limited hours or unlimited hours. The rest of the work were categorized as Risk, Marketing, Office or Development. All employees were located in the same office building. Fixed hours, limited hours and unlimited hours were further divided to different project by project ID.

Because office work shows up as a single project where everybody has reported working hours, there is a risk that the resulting network shows that all the actors in the network are connected through this project[1]. This could also result in that those people who solely report working hours for office work can become overly central in one-mode presentation. In some way it seems to be right, as the common sense suggests that people in administrative role in the office interact with many people and are central in communication compared to more focused work of a person in a specialist role. On the other hand, people participating

[1] Unequal distributions may mask the underlying tendency to interact through different types of projects. If the interest was the effect of project type to the occurrence of ties, then the analysis should take in to account the confound effect of different sizes of the projects (see Borgatti et al. 2013, p.238)

many projects interact with different people and can be regarded as central in communication networks. Those actors who participate in many projects have a higher probability of participating in the same project with other actors, and these actors also tend to become more central. On the other hand, dichotomization may remove the ties connecting these actors if their total work hours in each of the projects are not large enough.

Figure 6 shows properties of the project data. Distributions of the reported hours are skewed: many hours are reported to only a few big projects and the remaining hours are shared between numerous smaller projects. Similarly, many projects have only a few participants but a small set of projects was reported work hours by many people.

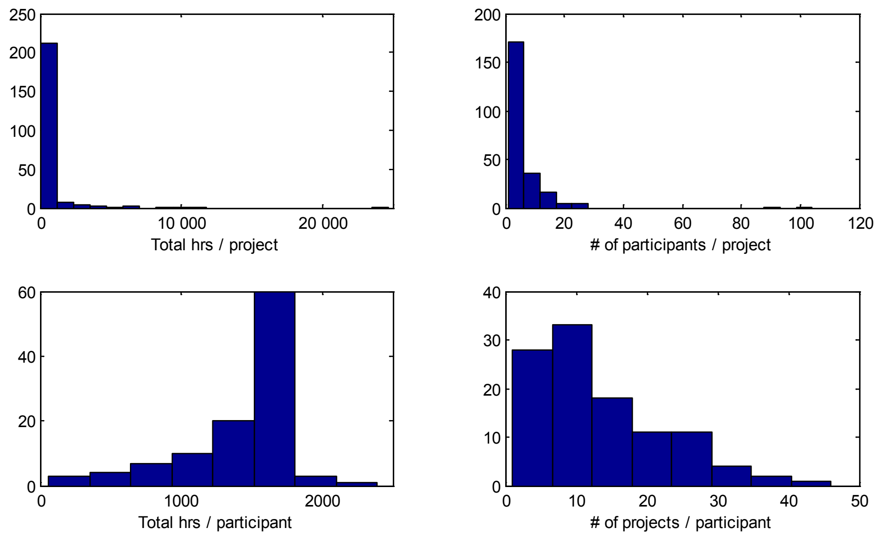


Figure 6 Properties of the project data: Distributions of total hours reported, project participants, total work hours and number of projects plotted against projects and employees. All distributions are profoundly skewed. What is the average project based on these data can be considerably different from what is commonly encountered project.

The routine network and the idea network are shown in

Figure 8, dichotomized at different threshold levels. The left column shows the routine network, and the right column shows the idea network, both with the monthly, weekly and daily communication rates. The networks dichotomized with the daily communication frequency show some typical structures of social networks: there are denser clusters, central actors, empty spaces spanned by actors and people on the periphery. The idea network is sparser; it has central actors but less visible dense groups. There were some isolates that were removed from the figure, but the giant component retains its shape and stays connected.

Figure 9 shows the project network at different threshold levels. The network graphs in the figure represent the project network dichotomized so that its density matches the densities of the self-reported routine network at the monthly, weekly and daily communication frequency.

The network in the middle row of the right column in Figure 9 shows the network near the maximum threshold level where one large component can be kept. When the cut-off value is increased from 500 hours, the giant component starts to break up into smaller pieces. The disappearance of the giant component is likened to a "transition change" in networks and gives a commonly used heuristic to choose dichotomization level. Another method is to set the threshold level where the giant component appears to grow at fastest rate. (Thomas & Blitzstein 2011b)

Social network analysis is often applied at node level, where various centrality measures are used to answer questions of the relative importance of the actors in the network (Wasserman & Faust 1994, p.169). One could argue that the dichotomization level should be set so that the relative importance of the different nodes in the network doesn't change as a result of dichotomization.

There are various generalization of centrality measures for valued networks, but they are not in common use. Typically weighted networks are dichotomized to binary networks. Maybe the simplest way to generalize degree centrality to weighted network is to take the sum of weights instead of the number ties (Barrat et al. 2004). Figure 7 shows how rank order of central actors differs as dichotomization level is changed. To produce the plot in the figure, steps used in Thomas and Blitzstein (2011b) were followed. First, the sum of the weights was calculated for each actor in the original valued network, and actors were ranked according to their scores. The process was then repeated for each different dichotomization level. A Rank disparity indicator was calculated for each dichotomization level using equation below and is shown in Figure 7.

$$D_{ab} = \frac{1}{N}\sum_{i}\frac{(R_{ai} - R_{bi})^2}{\sqrt{R_{ai}R_{bi}}}$$

Discrepancy was smallest at dichotomization level of around 50—100 hours. This suggests that dichotomization around this level retains some characteristics between the original valued network and the dichotomized network –at least if compared using Barrat's centrality measure. Using different centralization measures would probably indicate different optimum (cf. Thomas & Blitzstein 2011b). The top-right corner network in Figure 9 represents the project network dichotomized at 50 hours. At this density level, the network seems to be too dense to reveal any visual cues of its structure.

As self-reported networks have natural dichotomization levels arising from the design of the questionnaire, the approach chosen here is to run the analyses so that the project network has been dichotomized to have an equal density to the self-reported routine network. The plot in Figure 9 shows that some information will be lost with high dichotomization levels.

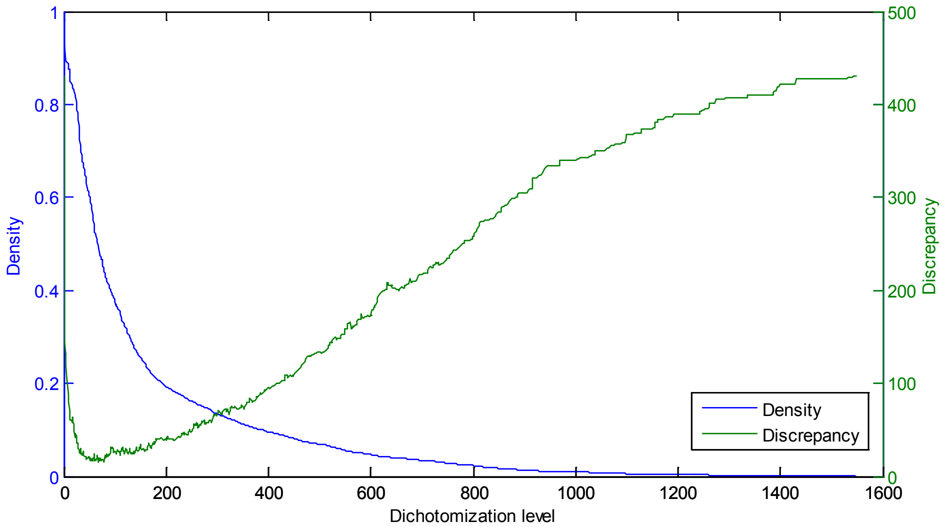


Figure 7 Discrepancy of Barrat's degree centrality measure between the original valued network and dichotomized networks at different dichotomization thresholds (green rising curve). Density of the network drops quickly as the threshold is increased (blue decreasing curve). The discrepancy of the rank orders shows global minimum at around 50-100 hours suggesting that dichotomization at this levels minimizes the loss of information.

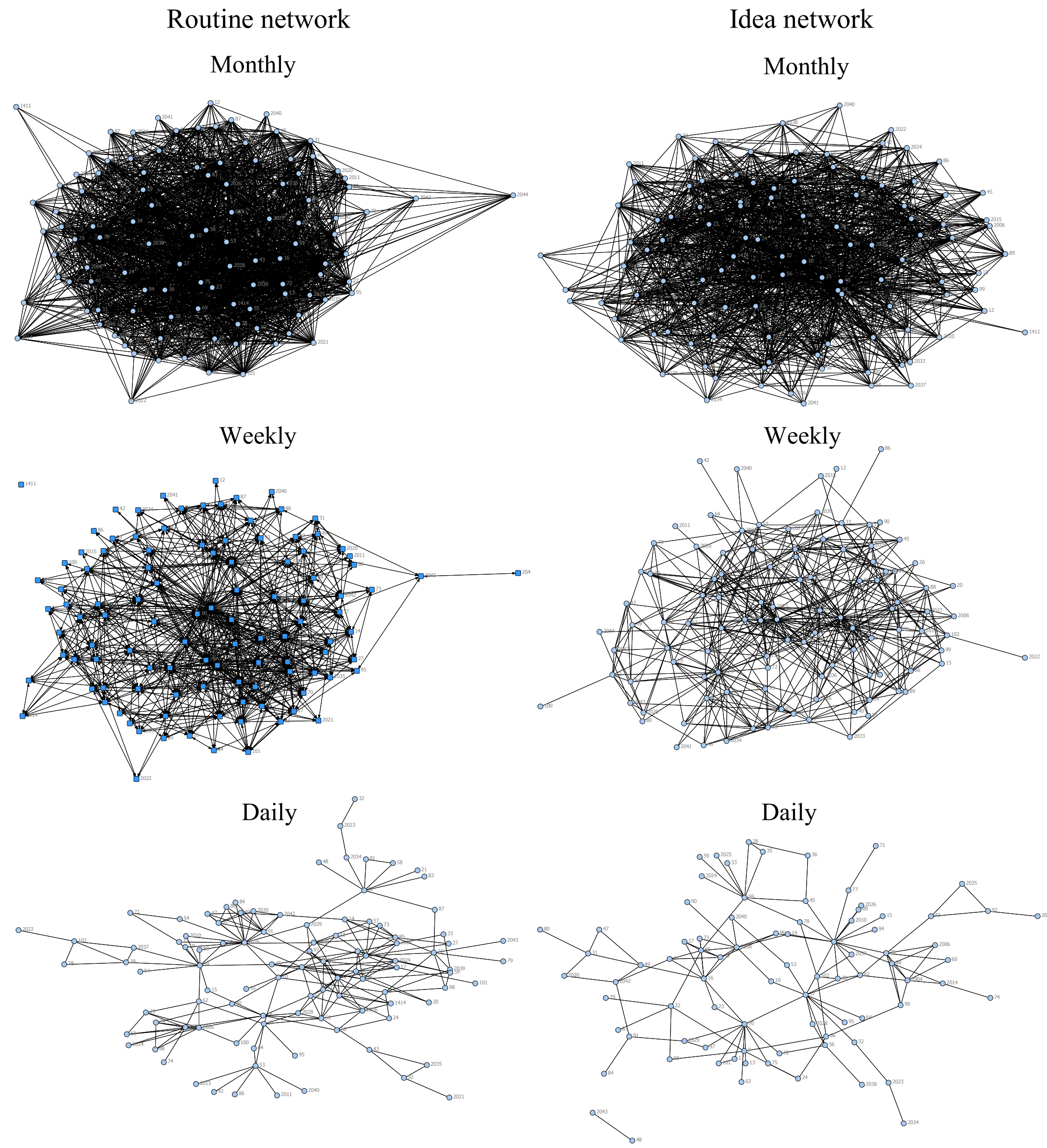


Figure 8. The routine and the idea networks at the monthly, weekly and daily frequency. Each dichotomization level brought up isolated nodes to graphs. These were removed from the visualization. Daily frequencies reveal visual structure in the two networks. Routine network has denser local clusters than the idea network. Both networks have actors connecting the denser parts of the networks together.

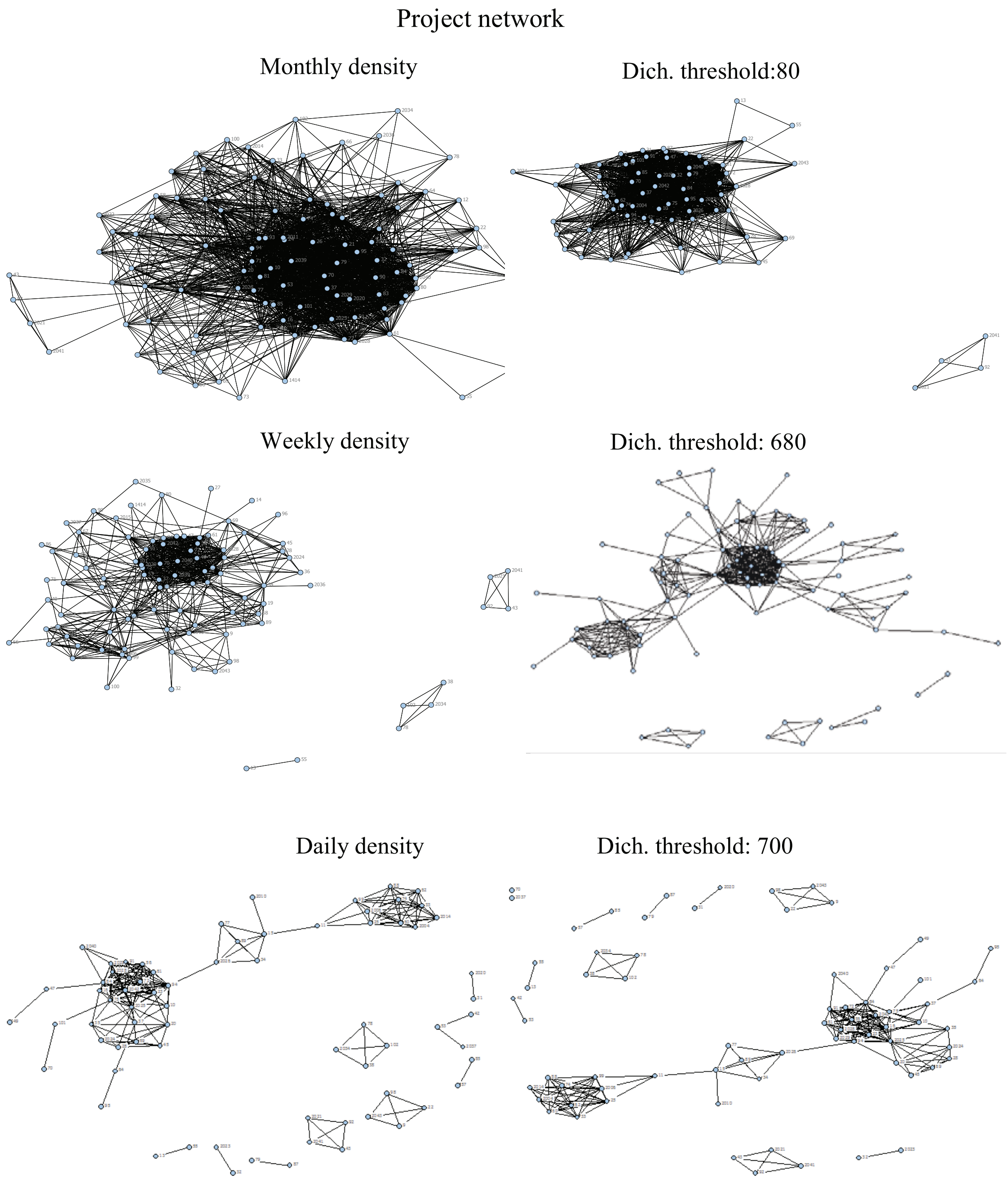


Figure 9.The project network dichotomized at different levels. The left column has project network dichotomized to correspond the density of routine network with monthly, weekly and daily frequency. On the right side, dichotomization level of 80 was the result of minimum discrepancy of Barrat's centrality measure. Dichotomizations at 500 and 700 show how the giant component changes shape between these dichotomization levels. The well-connected component straightens up to a long string, increasing betweenness of some actors and moving some actors to periphery.

5 Results

This section shows the results of the network analyses performed for the routine, idea and project networks. First, the networks were compared on whole network level using descriptive statistics. Next, networks were compared on the dyadic level by correlating their adjacency matrix representations and by comparing actor proximities by using multi-dimensional scaling or MDS. Fourth, networks were compared on the sub-group level: how many cliques the networks have, do they share the same actors in the cliques and is there observable core-periphery structure. Fifth, actors were ranked by the four most common centrality measures in order to see if the most central actors are the same in each network. The results of each level of the analysis are described below and further summarized and discussed in chapter 6: Discussion.

5.1 Comparison by descriptive statistics

The network matrices constructed from the survey and the project work hour data were matched to contain 108 actors. The actual number of respondents who completed all the questions in the survey was 77, but the actors who were mentioned by other network members were also included in the analysis. This way possibly important missing actors are not left out of the network structure because it is not unusual if important actors do not have time to answer questionnaires (Kossinets 2006; Borgatti et al. 2013).

Descriptive statistics of the networks are shown in tables 1 and 2.

Table 1 shows that the routine network is denser in every dichotomization level than the idea network. The project network was dichotomized to have the same density as the routine network.

Average degree shows that in average, routine work related matters were discussed with 38 people and with 3.9 persons every day. Ideas are shared with fewer people than routine related matters: In average, ideas were discussed with 31 different people and with 2.4 people daily. Project participants are collaborating nearly with the whole organization. Average degree measures in the project network are of the similar magnitude to that of the routine network because of the dichotomization decisions. The number of connections people can maintain reduces as communication frequency is increased.

Table 1. Descriptive statistics of the self-reported networks with different dichotomization levels

Routine network	**Monthly**	**Weekly**	**Daily**
No. of people	109	109	109
No. of ties	2079	760	210
No. of components	1	2	10
Density	0,35	0,13	0,04
Avg. Degree	38,15	13,94	3,85
Clustering	0.642 / 0.526	0.539 / 0.333	0.496 / 0.307
Centralization	0,61	0,81	0,12
Diameter	3	5	9
Idea network			
No. of people	109	109	109
No. of ties	1693	510	129
No. of components	1	2	28
Density	0,29	0,09	0,02
Avg. Degree	31,06	9,36	2,37
Clustering	0.596 / 0.471	0.454 / 0.282	0.364 / 0.200
Centralization	0,68	0,34	0,13
Diameter	3	5	10

Table 2. Descriptive statistics of the network constructed from the reported working hours

Project network	**Min. discrepancy[1]**	**Monthly**	**Weekly**	**Daily**
Dich. threshold	80	111	313	688
No. of people	108	108	108	108
No. of ties	3484	2019	748	209
Components	1	4	11	48
Density	0,603	0,349	0,129	0,036
Avg. Degree	64,519	37,4	13,9	3,8
Clustering	0.881/0.859	0.800/0.745	0.760/0.729	0.817/0.808
Centralization	0,300	0,377	0,230	0,135
Diameter	4	4	4	8

1) Min. discrepancy level was set to minimize the difference in the rank order of Barrat's degree measure between original and weighted networks.

The tables above contain also average clustering coefficient values. The second clustering coefficient value is a weighted clustering coefficient weighted by nodes degree. The weighted clustering coefficient is more suitable for comparison across networks with different densities as it weighs the actors by the sizes of their neighborhoods (Hanneman & Riddle 2005). The clustering coefficient shows that project network is more clustered than routine and ideas network. It is also more fragmented with a higher number of components.

Overall cluster density is much higher than the graph density, suggesting that networks demonstrate clustering.

When the project hour data was projected onto one-mode network, people working on the same project got mutual connections and formed fully connected local clusters. The dichotomization process done to the project network removed some of the low-intensity connections within and between clusters. As can be seen from the number of the components, some of the clusters were tied together with weak ties and these ties were cut as dichotomization level was increased.

The centralization measures in Table 1 and Table 2 decreased as networks were examined at higher dichotomization thresholds. The centralization of the networks at the daily density level is relatively low across all the three networks: around 13 % compared to a perfectly centralized star-shaped network. This indicates that there were some actors who have many weak connections, but as the threshold is increased, these connections are filtered out, and the networks became less centralized.

The diameter of the network is the length of the largest geodesic distance in the network or infinite if the network is fragmented (Wasserman & Faust 1994, p.112). UCINET sets the diameter to be the geodesic of the largest component. Diameter increases with dichotomization level. The diameters are low in all of the three networks as can be seen from fully connected monthly communication frequencies: it takes three steps for information to spread from one end of the network to the other.

Generally social networks are not uniformly dense. Actors may cluster to form denser areas as shown by appearing of components in the three networks (Kilduff et al. 2008). The background data of the network actors were available allowing sectioning of the overall density measure. Table 3 below shows how weekly density varies in the routines network in different organizational units (cf. Cross et al. 2002). Similar pattern was visible with different dichotomization levels in the other two networks as well (see Appendix D). Density seems to be higher inside the organizational units than between units.

Table 3. Densities between organizational units in communication of routines with weekly or higher frequency

Unit	1	2	3	4	5	6	7	8	9	10
1	0,33	0,11	0,08	0,15	0,15	0,04	0,05	0,05	0,08	0,00
2		0,27	0,08	0,13	0,17	0,08	0,08	0,03	0,06	0,00
3			0,80	0,13	0,40	0,07	0,05	0,09	0,05	0,00
4				0,29	0,23	0,09	0,07	0,03	0,05	0,03
5					0,44	0,22	0,12	0,25	0,18	0,00
6						0,52	0,02	0,02	0,13	0,04
7							0,16	0,05	0,02	0,00
8								0,67	0,13	0,06
9									0,50	0,04
10										1,00

Overall, discussion of routine work and ideas form well-connected networks where each member in the organization can be reached. On daily communication level, the networks separated to smaller components and one large component. All the networks show higher densities within organizational units than between different units, suggesting cohesion where the organizational unit is a factor. The within group densities of the idea network indicate that discussion of ideas does not happen as tightly within one organizational unit compared to the discussion of routine work or the co-occurrences in projects.

The project network differs from the routine and the idea network in the sense that it has the highest amount of clustering, and it also fragments into many more components than the two other networks. The routine network is more clustered as well as denser than the idea network on all dichotomization levels. Differences in descriptive statistics are summarized in table 4.

Table 4. Summary of the differences of the three networks based on descriptive statistics.

Routine network	Project network	Idea network
Fragments lightly when limited to stronger ties.	Fragments strongly when limited to stronger ties.	Fragments moderately when limited to stronger ties.
Routines discussed with 38 persons in average monthly and 4 persons daily.	Average number of people co-occurring in projects varies from 38 persons to 4 as dichotomization level is increased.	Ideas discussed with 31 persons in average monthly and 2 persons daily.
Centralization is high in weekly communication (81 %) and low in daily communication (12 %).	Centralization is generally low: varies from 38 % to 13% as dichotomization level is increased.	The idea network shows less centralization than routines network: Centralization is highest, 68 % monthly and goes down to 12 % in daily communication.
Network has small diameter of 3 steps, but regular daily communication forms a core with 9 steps diameter.	Network is not connected. The largest component has a diameter of 4 steps which increases to 8 steps in daily work.	Network has small diameter of 3 steps, but regular daily communication forms a core with 10 steps diameter
Moderately clustered	Highly clustered	Less clustered than routines network

5.2 Correlations between networks

One way to see similarities and dissimilarities between networks is to calculate Pearson correlation co-efficient for pairs of matrices. Correlation in this case measures the association between presences of network ties. In order to measure the correlation, the network matrices had to be sorted to have the rows in the same order. When needed, some actors were removed from the data to make the matrices same size.

Figure 10 shows Pearson correlation measure[2] calculated for the project network and for the routine and idea networks. The plots show that the correlation between the three networks depends on the decision of how project network is dichotomized. As explained in section 4.11.3, the project network was dichotomized so that its density corresponds to that

[2] Pearson's r was used extensively in literature to measure association between networks (cf. Hanneman & Riddle 2005) compared to for example Spearman's ρ which in social sciences is often used to measures association between variables measures in ordinal scale and in some network comparisons (Johnson et al. 2012 for example)

of the routine network. This implied dichotomization levels of 111, 313 and 688 hours to make its density correspond to monthly, weekly and daily routine communication.

Figure 10 shows that the ties in the project network correlate generally more with the routine network than the idea network. As dichotomization level is increased, the correlations become equal as the ties that make the network different start to disappear.

Table 5 shows the results of correlating the project network with the routine and the idea networks when dichotomized at the levels mentioned above. It is not possible to calculate confidence interval for the correlation coefficient, but it was possible to calculate significance levels by permutation-based estimation (see section 4.8.1 for a description of the method). All correlations were statistically significant with $p < 0.05$ as calculated using quadratic assignment procedure with 20000 runs.

The ideas network and the routine network are highly correlated with each other. At weekly and daily frequency of discussions, the project network ties correlate clearly more with the routine network ties. The correlation between the project network and the other networks is quite small in any case and attains the maximum value of 0.35.

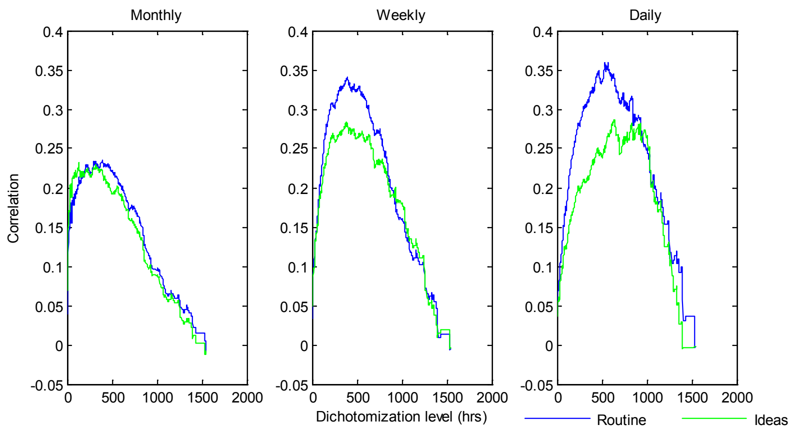


Figure 10. The correlation between project network and the routine and ideas networks with different communication frequencies.

Table 5 Correlations of occurrences of ties at different dichotomization thresholds

	Monthly			Weekly			Daily		
	I	**P**	**R**	**I**	**P**	**R**	**I**	**P**	**R**
Idea network	1	0,213	0,696	1	0,269	0,611	1	0,263	0,599
Project network		1	0,191		1	0,327		1	0,312
Routine network			1			1			1

All correlations were statistically significant with *p<<0,001*. Significance was measured using Quadratic Assignment Procedure (as implemented in UCINET 6), where *p*-value is the proportion of times that correlation in 10 000 permutations was larger than or equal to the observed value.

5.3 Comparison by actor proximities

Multi-dimensional scaling offers a visual way of comparing similarities of different networks. MDS assigns the original data observations new points in two-dimensional space in such a way that the distances between the points in a two-dimensional space correspond to the proximities in the original data. (Knoke & Yang 2008, pp.82–85)

The routine, idea and project networks are weighted networks where weights represent how frequent communication is or how many hours the actors could have collaborated in the same projects. MDS can be applied to the original valued adjacency matrices after the matrices are symmetrized. Symmetrization guarantees that proximity of an observation A to B is the same as proximity of observation B to A.

The values in the three networks can be loosely interpreted as proximity data. More frequent communication knits actors more tightly to their counterparts and in this sense nearer to them. Proximities do not satisfy most of the properties of distance functions: Here the proximities are not additive, and zero-proximities do not imply that actors are coincident in the network. Actually, zeros in the adjacency matrix mean that actors are isolates. Finally, proximities do not necessarily satisfy the triangle inequality because using a mediator can be faster route to pass a message than a direct contact. For these reasons, Euclidean or other distances cannot be used, but non-metric MDS with iterative algorithms can be useful. Non-metric MDS is not too sensitive of the absolute differences of the proximities but tries to keep the order of proximities the same.

MDS algorithm was first run for routine network, which is a valued network on an ordinal scale from 0 to 3. Analysis of the routine network was first performed in an exploratory way with no particular restrictions onto the MDS starting configuration. Ideas network and project network were then scaled using the results of the routine network as a starting configuration.

Project co-occurrences are ratio data in hours, ranging from 0 to 1634 hours. For the project network data, MDS tended to give degenerate solutions with lots of nodes lying on different corners on top of each other. Project network was pruned until the MDS algorithm could generate a valid presentation of the network. The outlier nodes that caused degenerate solutions were removed (ID = 15, 55, 13, 1411). The outlier nodes were the CEO and some employees from the financial or IT- support functions whose work hour profiles differed from the other actors. Office assistants stood out of the MDS analysis as well. They reported working hours mostly or only as office hours. It is possible that assistants help a large number of other actors and are highly central in the network.

Figure 11 below shows the results of running non-metric MDS algorithm as implemented In UCINET 6. The MDS solutions have relatively high stress values of 0.265, 0.146 and 0.167 for the project, ideas and routine networks respectively. Because of this, the nodes in the central area of the MDS plot are best treated as arbitrarily positioned. The error in the periphery however should be relatively smaller and is used here to compare the results (Borgatti et al. 2013, p.91).

The project network produced visually different MDS presentation from the routine and idea networks. The MDS plot of the project network has a dense center with isolated nodes surrounding it whereas the MDS of the routine and idea networks produced maps with nodes scattered evenly to a wider area. Nodes that are located in the edge of the plot in the MDS of the project network and the routine network have been marked on the plots. Although there are differences how the MDS results look like, the nodes located far from the dense center in the project network seem to be also far from the center in the MDS of the routine network. Some pairs of nodes that are far from each other in the project network are far from each other also in the routine network. On the other hand, some node such as node 92 is in the edge of the project map but located in the center in the routine and idea networks.

The data of the ideas and routine networks were collected with the same questionnaire and their MDS representations look very similar to each other. Many of the nodes that are located on the edge of the MDS of the routine network are also located at the edge in the MDS of the idea network. There are groups of nodes that are located near each other in both of the maps. These groups are highlighted by hand with light blue background color. Generally, points in both of the maps are scattered across the space without a clear pattern, and there are no recognizable axes.

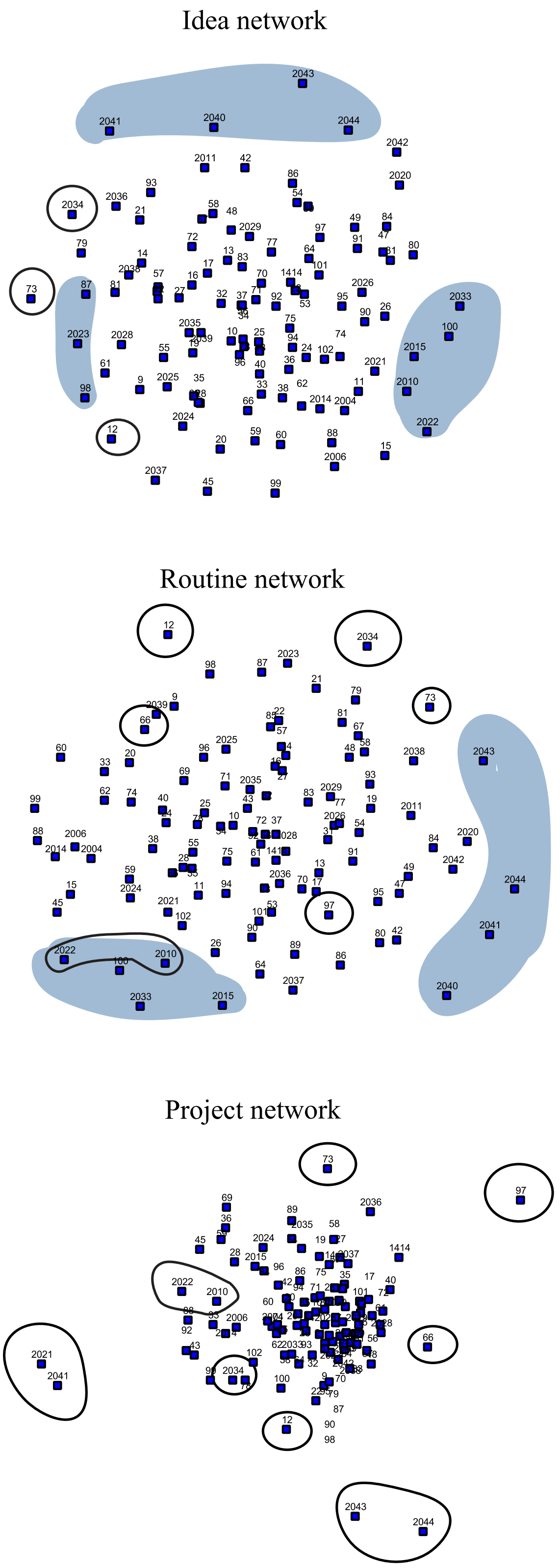


Figure 11. Non-metric MDS of the idea network (top), the routine network (middle) and the project network (bottom). The MDS of the project network shows a dense center and actors in periphery. The routine and idea networks show similar plots. The boundaries are drawn by hand to aid comparison of groups of nodes in the project and the routine networks. The blue markings highlight groups of nodes that occur in the idea and the routine networks.

5.4 Comparison on the sub-group-level

Above, the networks were compared on the whole network level by using various descriptive statistics and on the dyadic level by correlating network ties and by visualizing actor proximities using MDS. Networks can also be compared on the sub-group level. Properties like the number of the different kinds of groups, their sizes and densities can help to describe the characteristics of the network as a whole. Some individuals are members of several groups and connect separate denser areas in the networks together. On the other hand, groups can explain and predict many of the whole network's behavior such as possible fights or separation into factions (Hanneman & Riddle 2005).

5.4.1 Clusters

The descriptive statistics in section 5.1 revealed that all three networks show clustering and that clustering was strongest in the project network.

The figure below shows a graph presentation of the project network dichotomized to the weekly density of the routine network. Various sizes of k-cores are highlighted in the graph with different colors. Some projects form large clusters of actors, which differs from the structure of the routine and idea networks.

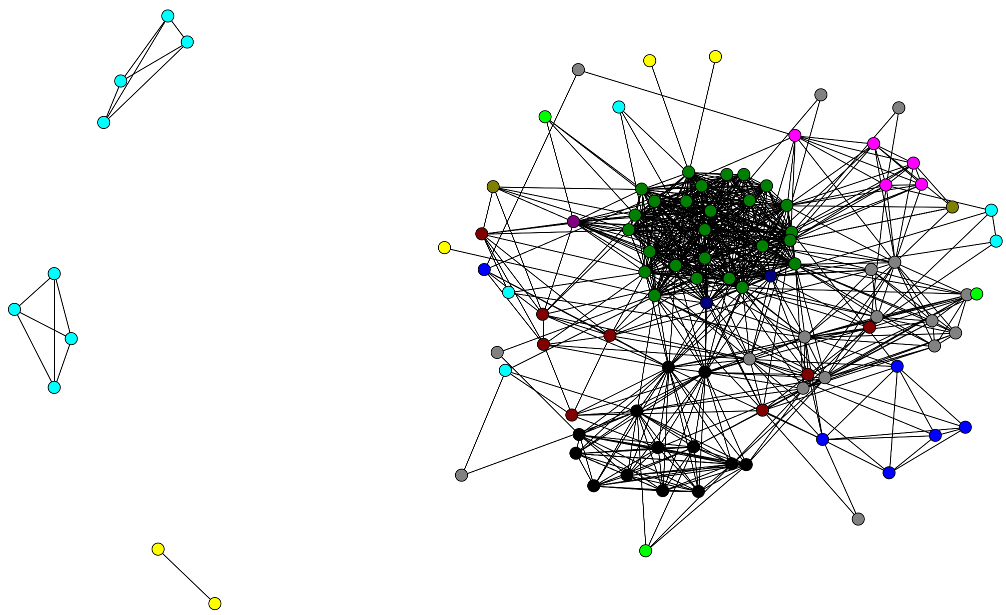


Figure 12. The project network shows large clusters of actors around projects of various sizes.

The large cluster probably arose because some projects were used to pool various tasks into categories that were not directly billed from the customer. Figure 13 shows the same project data after the data was partitioned to include only those projects that were classified as billable customer projects. The network still clusters, but around different projects after removing office, development and marketing work.

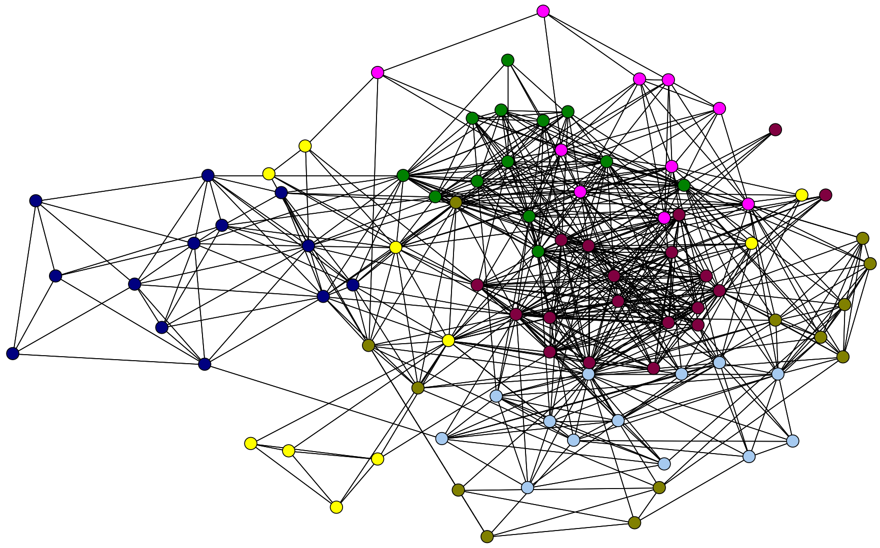


Figure 13. The project network after removing office, development and marketing hours.

5.4.2 Cliques

Cliques are maximum possible set of actors which have all possible ties present (Hanneman & Riddle 2005). They represent strong cohesion in social relations and at the same are probably visible in all the three networks.

Cliques overlap easily, and even small network can produce hundreds of overlapping cliques. An algorithm for finding cliques was run as implemented in UCINET for all the three networks. The networks were dichotomized to correspond to the density of daily communication to limit the number of cliques.

The diagonal entries in Table 6 show the number of cliques with three or more members in each network. The routine network had 53 overlapping cliques and largest cliques had four members. The idea network had 22 cliques, and the project network had 15 cliques with three or more members. The off-diagonal entries in Table 6 show how many shared cliques the networks have. Eight of the cliques found in routine network were also in ideas network while fourteen of the cliques in ideas network were also in routine network. The numbers differ because routine network included 23 four-person cliques, and many three person cliques of the ideas network were included in these. The routine and the idea network has largest cliques of four persons, and the project network had largest clique of 12 persons.

The results show that many of the cliques that can be seen in daily discussion of ideas and routines are also in the project network. The project network includes more of the cliques of routine work discussion than cliques of discussing ideas. Similarly, many of the cliques of the idea network can be seen in routine work related discussions. Cliques between the routine and the idea network have more in common. Interestingly, none of the cliques in the project network are visible in routine network or idea networks. Project network cliques are too large to be seen in the other two networks.

5.4.3 Active Clique members

Cliques can be used to measure association between actors. Actors who are members of many cliques and co-occur in the cliques at the same time can be seen to be close to each other. These co-occurrences can be collected to a clique-co-membership matrix, where the element $A(i,j)$ indicates how many times actor i is in the same clique with actor j. This kind of proximity data can be used as an input for hierarchical clustering method to produce a list of non-overlapping groups. Hierarchical clustering results can then be used to easily spot which actors do not participate in any groups and which actors are most actively participating in many groups. (Borgatti et al. 2013, pp.185–186).

Table 7 was created by creating co-occurrences matrix of clique memberships and applying hierarchical clustering to it. The actors listed in Table 7 are those pairs and groups of actors who were the most co-occurring actors in different overlapping cliques. The number in parenthesis shows how many different groups the actors were classified by hierarchical clustering.

The group hierarchy was lower in project and idea networks. Mostly different actors were the most active participants in these three networks. Actor 2006 was a member of many clusters in the project network and the routine network. Actors 16, 85 and 2004 were members in many cliques in the routine and ideas networks.

Table 6. The number of shared cliques of size three or larger between networks.

	Routine	Ideas	Project
Routine	53	8	0
Ideas	14	22	3
Project	1	1	35

Table 7. The most active pairs of actors participating in many cliques. The table shows the IDs of the actors and in parentheses, the number of different groups recognized by hierarchical clustering of clique membership matrix.

Routine	Ideas	Project
2026, 91 (11)	85, 16 (4)	23, 75 (4)
2004,2006 (5)	2004, 25 (4)	48, 67 (3)
16, 85 (5)	34, 62 (2)	45, 59 (3)
		16, 91 (3)

5.4.4 Core-Periphery Structure

The core/periphery partitioning is based on the idea that the network can be sectioned into two parts: relatively dense core with people being "in" and sparser periphery of outsiders. Borgatti and Everett (2000) have developed discrete and continuous core/periphery models and implemented them in UCINET to perform this kind of partitioning. Here, a continuous model is applied as it offers an additional advantage of producing proximity matrix for visualization.

The continuous version of the core/periphery model is based on correlating the original data matrix and a constructed matrix Δ where elements are products $\delta_{ij} = c_i c_j$. The number c is a *degree coreness* value of each actor. Elements in Δ get large values when the pair of actors c_i and c_j both have high coreness values. The objective is to maximize the correlation between the data matrix and Δ-matrix by changing values of c. This can be achieved numerically by using a suitable optimization algorithm. (Borgatti & Everett 2000)

Coreness measure resembles centrality measures. Instead of maximizing correlation, it is also possible to minimize the sum of squared differences between the two matrices. In this case the vector $\boldsymbol{c}$ is principal eigenvector and coreness measures coincide with eigenvector centrality (Borgatti & Everett 2000).

Figure 14 below shows the three networks with the core and periphery nodes colored. Coreness measures were calculated for each node using UCINET. Core/periphery-algorithm was run using correlation as a measure of fit and by running 20000 iterations. The routine in UCINET also included rules to classify nodes to core or periphery. The routine returned a correlation measure which compared the solution to idealized block model, where the suggested core nodes have a value of one and periphery nodes have a value of zero. Correlations of the idea, routine and project networks with an idealized block model were 0.478, 0.518 and 0.586 respectively. The coreness values were used as an input for MDS routine to produce coordinates for visualization of the networks.

The project network has a more pronounced core than the idea and routine networks. Core/periphery routine in UCINET recognized a core size of 29 actors for the idea network, 35 actors for the routine network and 25 actors for the project network. Many of the actors seem to be central in other networks as well. Project and routine network shared 16 actors in the core, ideas and routine network 23 actors and ideas and project network 17 actors in the core. Project and routine networks demonstrated slightly more core/periphery structure than ideas network.

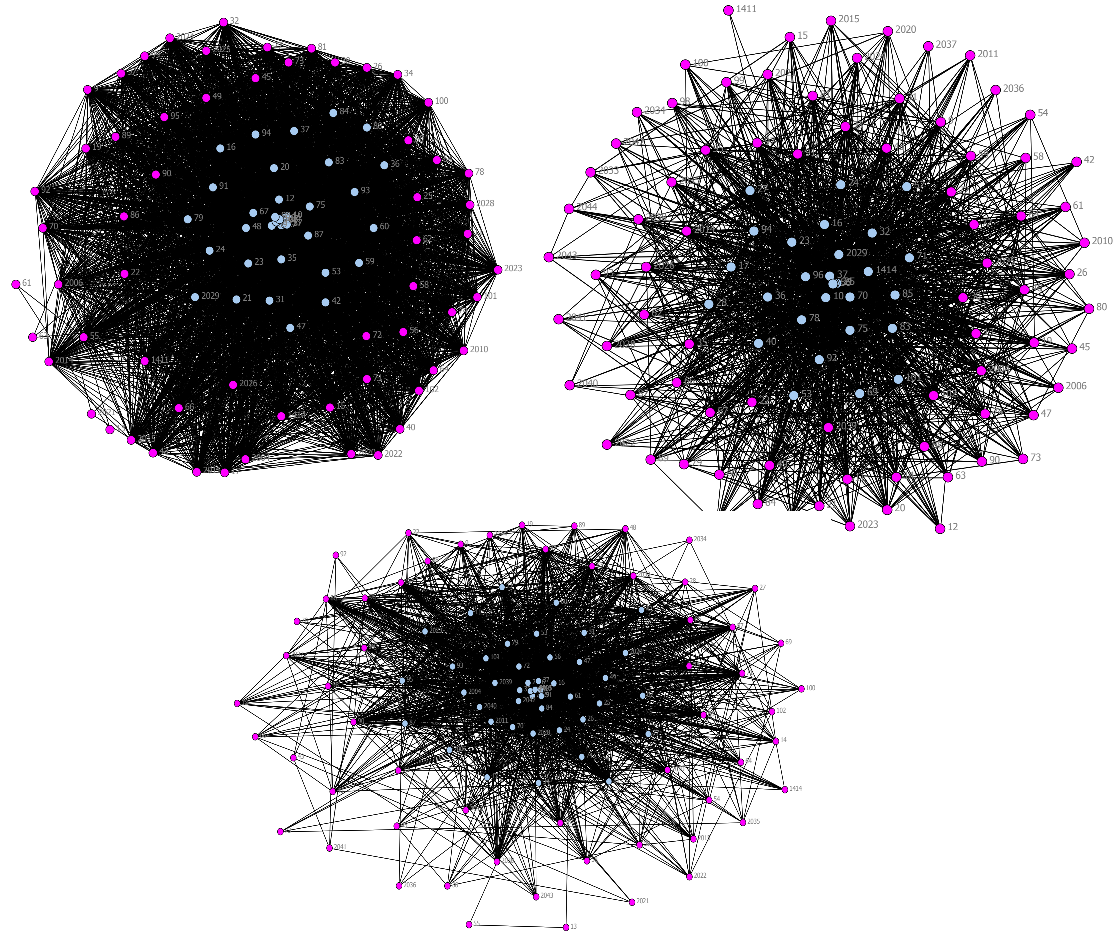


Figure 14. MDS visualization of coreness values of actors in the routine (top left), idea (top right) and project networks (bottom). All of the networks show signs core-periphery-structure.

5.5 Actor-level comparison

Actor level analyses focus on actors' positions in the social structure. Centrality measures are used to identify who occupy central positions in the network. Central actors tend to be the most visible; they know many people and are known by many people. There are multiple measures of centrality which are either designed to improve the simple degree measure or focus on different dimensions in the notion of centrality. (Prell 2012, p.96)

The graphs of the routine and idea networks in

Figure 8 showed that there are some degree of clustering in them, and some actors are more central than others. The graph visualization of the project network in Figure 9 showed that the actors in this network tend to take places in symmetrical cliques where members have more equal roles.

To identify the most central actors in the three different networks, four popular centrality measures were calculated: Freeman degree centrality, eigenvector centrality, betweenness centrality and closeness centrality. These centrality measures are defined for dichotomized networks only.

Different centrality measures were discussed in detail in section 4.4. Degree centrality shows the number of direct contacts an actor has. Eigenvector centrality weighs each actor's degree measure by how connected the actor's alters are. Betweenness centrality measures the extent that the actor is located on the other actors' geodesic paths. Closeness centrality on the other indicates how near the actor is to the other actors in the network. Some social network analysis packages give farness in the output which is reciprocal of closeness. (Prell 2012, pp.95–109)

In order to calculate the different centrality measures, the same dichotomization rules were used as in previous analyses. The centrality measures were calculated using the data of daily frequency of communications. The project network was dichotomized so that its density corresponds to the routine network at the daily density.

Table 8 shows the ten most central actors in each network measured with four different centrality measures. All the networks have mostly the same actors in the most central positions regardless of the centrality measure used. Picture changes somewhat if the whole column of a centrality measures is considered. Table 9 shows correlations between the centrality scores in each network. Centralities in the routine and the idea networks are moderately correlated ($r \approx 0.2 - 0.7$), except for betweenness centrality. Centralities in the project network correlate more with the routine network ($r \approx 0.2 - 0.3$) than the idea network ($r \approx 0.1 - 0.2$). The three networks show no correlation in their Betweenness centrality scores.

A large number of centrality measures are easier to digest with the help of visualization. Figure 15 shows overview of weekly level centrality measures for each of the networks using correspondence analysis. Before running the correspondence analysis, the centrality measures were normalized to between zero and one, and nodes with zero degree were removed. Correspondence analysis places different centrality measures to different corners in the resulting plots. Degree centrality measure is located in the center area of the plot indicating that degree centrality is related to the three other measures (Borgatti et al. 2013, p.95).

Correspondence analysis produced quite similar scatter clouds for the idea and the routine network. The plots suggest that there are similar network positions for actors in these two networks. Even though different actors are differently central in these networks, dots are scattered similarly between the corners of the plot. The plots of the routine and the idea networks show a diagonal pattern reaching from closeness centrality toward eigenvector centrality.

The nodes in the plot of the project network form two separated clouds centered on eigenvector centrality and closeness centrality. The project network has fewer actors scoring high in betweenness centrality than the routine and the idea network. The crowded area near the closeness centrality dot may come from the fact that the project network has more components than the routine and idea networks. UCINET calculated closeness measures for disconnected actors as the highest closeness measure plus one. The dots are also near each other in the proximity of the eigenvector centrality and closeness centrality.

Table 8. Ten most central actors in each network by different centrality measures

	Degree		Eigenvector		Betweenness		Closeness	
	ID	**Measure**	**ID**	**Measure**	**ID**	**Measure**	**ID**	**Measure**
Idea network	34	45	34	0,307	34	971	37	189
	37	36	37	0,282	37	880	34	194
	25	33	25	0,232	2029	755	56	207
	2029	30	56	0,212	25	434	25	208
	85	23	2029	0,197	85	357	2029	209
	56	23	23	0,190	56	301	23	209
	23	20	1414	0,175	23	250	1414	213
	10	19	10	0,169	2038	223	75	219
	96	18	17	0,162	96	212	70	221
	70	18	96	0,160	70	194	96	222
Routine network	56	100	56	0,382	56	2808	56	121
	2026	33	75	0,200	91	172	75	188
	75	33	23	0,195	2026	169	23	189
	23	32	1414	0,183	36	148	2026	190
	10	30	37	0,175	75	137	91	192
	36	30	91	0,173	34	109	37	192
	37	30	72	0,167	2043	108	36	193
	91	30	16	0,166	37	107	10	193
	1414	28	10	0,162	53	97	1414	193
	16	27	35	0,160	10	93	53	195
Project network	2026	38	2026	0,193	2011	507	2025	232
	20	36	37	0,193	23	312	53	235
	2025	36	2039	0,193	26	288	20	235
	2039	36	94	0,192	79	268	2029	235
	37	35	20	0,191	53	244	37	235
	94	34	2029	0,191	2026	207	23	236
	2029	34	2025	0,190	2025	200	24	239
	53	34	75	0,189	20	188	70	240
	35	34	70	0,189	93	188	2039	241
	23	34	31	0,187	16	177	94	241

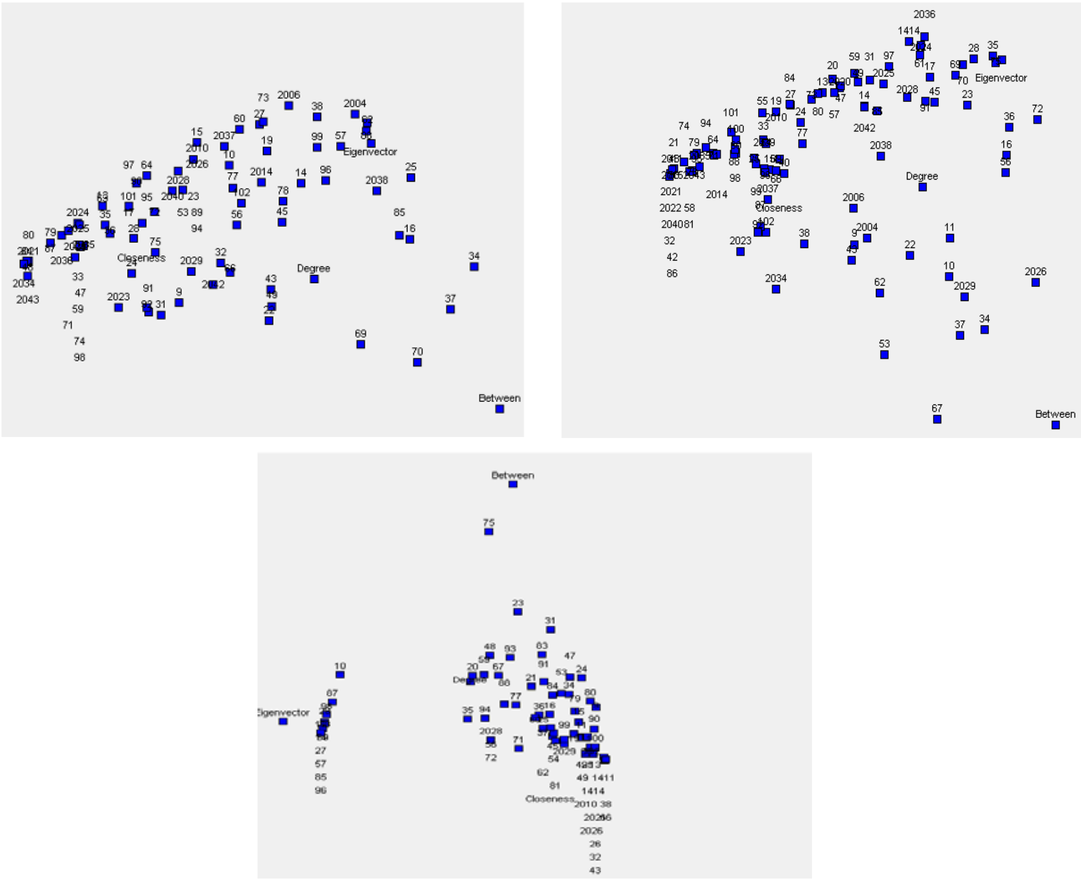


Figure 15. Four different centralities visualized using correspondence analysis for the idea network (top left), the routine network (top right) and the project network (bottom). Actors' centralities show similar patterns in the routine and the idea networks. Centralities in the project network form two separated areas near eigenvector centrality and closeness centrality.

Table 9. Correlations between centrality measures in each network. The centrality scores between the routine and the idea networks are moderately correlated. Scores are more correlated between the project and routine networks than project network and idea network.

Degree

	Routine	*Idea*	*Project*
Routine	1		
Idea	0.545	1	
Project	0.381	0.269	1

Closeness

	Routine	*Idea*	*Project*
Routine	1		
Idea	0.627	1	
Project	0.347	0.301	1

Eigenvector

	Routine	*Idea*	*Project*
Routine	1		
Idea	0.678	1	
Project	0.507	0.261	1

Betweenness

	Routine	*Idea*	*Project*
Routine	1		
Idea	0.198	1	
Project	-0.019	0.155	1

6 Discussion

The research question in this thesis concerned what structural characteristics do networks of the project participation and discussing routine work and ideas have. The theory review and the research question lead to the hypothesis that the network of discussion of routines and the network of project participation are structurally similar to each other while these two networks are structurally different to the network of discussion of ideas. To answer the research question, analysis results from the previous section are combined and summarized to describe the characteristics of each of the networks. The findings are then used to find support or falsify the research hypotheses.

6.1 Summary of the network characteristics

The graph visualizations in

Figure 8 and Figure 9 give already a good overview of the similarities and differences of the tree networks. The differences are easily seen on the dichotomization level corresponding to the daily communication density. The routine network is characterized by denser areas and well-connected central actors. The idea network looks very similar, but is sparser, and there are less visible clusters. The clusters and the central nodes in the idea network seem to be located in the same places as in the routine network (see Appendix C for larger picture). The project participation network has a dense core which fragments to separate components when the dichotomization level is increased. Large projects are visible in the network as dense clusters. The routine and the idea networks have peripheral nodes which are connected to the rest of the network with one or two ties. The peripheral nodes in the project network look different: The central component is surrounded by disconnected components and the nodes on the boundary of the large component are connected by multiple ties.

The network level descriptive statistics in Table 1 and Table 2 show many features that were visible already in the visualizations. Employees discussed on routine work with a higher number of alters as they did on ideas. Networks at daily communication frequency are sparse: maintaining daily communication requires time investment from the actors (Hansen 2001). All the networks were denser within organizational units than between units (see Appendix D). Participation to projects and discussions on routines formed more clustered networks than sharing of ideas. Overall centrality in the routine and the idea networks fell as dichotomization level was increased, and both networks were generally more centralized than project network. Centralization in the project network did not change much between different dichotomization levels. The project network had the shortest diameter; the routine network had a longer diameter and the idea network had the longest diameter.

On the dyadic level, Table 5 showed that occurrences of network ties were highly correlated between the routine and the idea networks and there was less correlation between these networks and the project network. Ties in the project network were more correlated with the ties in the routine network than with the ties in the idea network. Betweenness centralities were uncorrelated.

Multi-dimensional scaling (Figure 11) allowed to visualize differences in tie strengths. The proximities of the project network formed a dense cloud which was surrounded by actors far away from the center. The shapes of the scatter clouds of the routine and the idea networks were quite similar. In both of the plots, some of the actors were located similarly either in the center of the map or far from the center. Some actors formed recognizable groups outside the center in the MDS of the routine and the idea networks (highlighted in the figure), but still many actors were located differently in both of these maps. This suggests that these two networks have some degree of structural similarity, but they are different kinds of networks. As the stress values of the results were quite high, not too much can be said about the locations of the center actors. The MDS map of the project network looks different to the routine and the idea networks, but there are some actors that were on the outer hemisphere also in the routine network. Figure 11 has circled marking to indicate some actors in both, the project and the routine networks. Similarities with these locations could indicate that the project network reflects some features of the underlying social structure that also appear in the routine network.

Cliques are the most cohesive groups in social networks, and they probably contain the strongest ties. All networks had a large number of cliques in the level corresponding to the daily communication density. The project network differed from the routine and the idea networks by having much larger cliques surrounding the projects. The largest clique in the project network had 12 persons when the network was dichotomized to correspond to the density of daily discussions. The number of cliques rose quickly as the networks were studied with weaker ties. The largest cliques in routine network had 5 persons and in the idea network the largest cliques had 4 persons. Over half of the cliques in the idea network also appeared in the routine network, but only one-sixth of the cliques in the routine network were also in the idea network. Cliques in the project network were too large to be seen in the two other networks, but many of the cliques in the routine network could be seen in the project network. The results show that the social circles people form in discussing the work related matters overlap, but also differ from how groups are formed around the projects in the organization. It seems to be more difficult to be included into the different social circles formed by discussion than to be a member of a project.

Mostly different actors were the most active clique members in the three networks. Hierarchical clustering did not reveal as complex clique hierarchy in the project network than in the routine network. The idea network seemed to have the smallest number of levels in group structure. Finally, all the networks showed some degree of core/periphery structure. The project network showed this characteristic the most, then the routine network and the idea network a bit less.

The actor level analysis with the daily communication density shows that the most central actors were central in all of the networks. Overall, high centralities in the idea network tend to be accompanied by high centralities in the routine network. The centrality measures in the routine network and the project network were more correlated than centralities in the idea network and the routine network. The actors in the project network were more similar to each other in closeness centrality and eigenvector centrality. This seems to indicate that actors in a project network are equally close relative to the whole network, and they are central if their neighbors are also central. Scatter clouds in the correspondence analysis results (Figure 15) show the differences of the large number of centrality measures on weekly communication. One similarity between all the networks is that points in the correspondence analysis tend to stay far from the betweenness centrality, suggesting that betweenness centrality wasn't significant in any of the networks.

The characteristics of each network on different levels of network analysis are summarized in Table 10.

Table 10. What structural characteristics do networks of project participation and discussing on routine work and ideas have? Summary of findings.

Network level	**Routine network**	**Idea network**	**Project network**
Whole network	Denser than the idea network, slightly centralized, shows clustering and empty space between clusters.	Sparser than the routine network and less clustered, slightly centralized, empty space between clusters.	The least centralized, heavily clustered, fragmented and shows connected giant component surrounded with floating components.
Dyadic level	Occurrences of ties correlate moderately with the idea network. MDS revealed groups of actors with similar location in MDS to the idea network and some individual points with similar location in the MDS of the project network.	Occurrences of ties correlated moderately with the routine network. MDS revealed groups of actors with structural similarities to the routine network.	Occurrences of ties slightly correlated with the routine and idea networks. MDS revealed individual actors who differ from the rest of the actors similarly in the routine network. MDS map differed from those of the routine and idea networks.
Subgroup level	High number of cliques from 5 to 7 persons. Shows slightly more core/periphery structure than in the ideas network.	Many cliques from 3 to 4 person but less than in the routine and the project network. Slightly recognizable core/periphery structure.	Not many, but large cliques up to 12 persons. More core/periphery structure than in the idea network.
Actor level	Centrality measures distribute evenly for closeness centrality and eigenvector centrality. Only few actors with high betweenness centrality.	Centrality measures distribute evenly for closeness centrality and eigenvector centrality. Only few actors with high betweenness centrality.	Centrality measures formed two groups around eigenvector centrality and closeness centrality. Degrees distribute more egalitarian way. Very few scored high in betweenness centrality.

6.2 Research question and research hypotheses

The research question and the theory review led to hypotheses about the expected differences in the networks.

HYPOTHESIS 1: The project co-occurrence network derived from the project work hour data is structurally more similar to the network of discussion of routine work than to the network of discussion of ideas.

Hypothesis H1 gets support from the analysis. The routine network and the project network showed higher clustering than the idea network. Multi-dimensional scaling showed more similarly positioned points between the project network and the routine network than between project network and idea network. Occurrences of ties showed more correlation between the routine and the project network than between the project and the idea network. The routine and the project network showed slightly more core/periphery structure than the idea network. Centrality measures of actors correlated more between the project network and the routine network than between the project network and the idea network.

HYPOTHESIS 2: Discussion of ideas forms structurally different network to the network of discussion of routine work and to the project co-occurrence network.

This hypothesis was partially supported. Discussion of ideas has a network structure that is more similar to the routine network than the project network. Visualizations of the network (

Figure 8) showed how denser areas and central actors are located nearly in the same locations in the graph. In every level of the network analysis, the idea network showed more similarities with the routine network than with the project network.

There were some differences between the idea and the routine networks. The idea network was less clustered than the routine and the project networks. The MDS map of the idea network had differences to the routine network, but some groups of actors were located similarly in both MDS maps. Overall, the idea network had more differences to the project network than the routine network had differences to the project network. Hypothesis 2 cannot be falsified, but it doesn't get strong support either. The idea network had some differences to the routine network, but from most parts it was very similar. The project network was different to both of the networks.

The figure below outlines the expected results and the results got. The sketch on the left side of Figure 16 shows how the project network was expected to be different in some dimension to the routine and the idea networks as the project network is derived from a data serving as a proxy for the underlying relation. The project network was expected to show more similarities with the routine network than the idea network. The results got are more in line with the right side sketch. The routine and the idea networks share a lot in common and are different to the project network, but there were less differences between the routine and the idea networks than differences between the idea and the project network. The sketch on the right side demonstrates how the routine network is in between the idea and the project network. The arrow lengths represent the amount of difference between the networks. The project network is drawn slightly aside of the others to represent the additional latent factors that explain its structure.

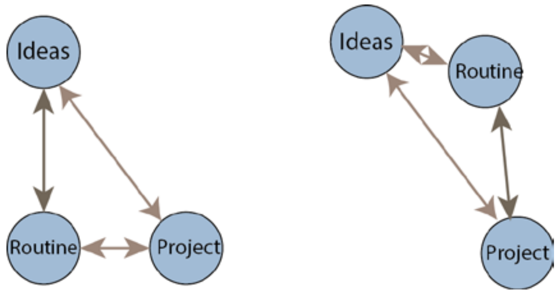


Figure 16. An outline of how the three networks were expected to be similar (left -side) and what results were got (right-side)

6.3 Possible reasons for the observed differences in the networks

According to the structural hole-theory, actors who are located near structural holes have vision advantage, and they are probably the ones to hear new ideas first (Burt 2004). Because generating ideas requires combining existing and new knowledge in new and creative way, actors at the structural holes should be among the ones recognized as good sources of information (Burt 2004; Perry-smith et al. 2006; Koestler 1989). Idea network was the sparsest of the three networks and has potential for structural holes. Sparse structure gives also more opportunities to bridge various parts together across the structural holes.

Idea network was least clustered of the three networks compared. Because only the ties with the most frequent interaction were studied, this indicates that the idea network had the least amount of cohesive clusters that could contain strong ties. As was discussed in section 2.1.3, cohesive groups tend to cause similarities in how their members think and behave and they are not optimal for idea generations (Burt 2004; Perry-Smith & Shalley 2003; Perry-smith et al. 2006). Having ties to various parts in organizations brings fresh ideas and adds different viewpoints to thinking (Burt 2002).

The questionnaire described ideas as “light bulb moments” suggesting novelty and creativity in ideas (see Appendix A for questionnaire). It is thus likely that respondents indeed meant discussion of novel ideas instead of, for example, small improvements in effectiveness in their routine work. According to the theoretical review in section 2, novel ideas usually come from discussions with people with different backgrounds.

The networks of project participation and discussion of routine work showed high clustering. This result was also expected as routine work is likely to be discussed with people one is working with in a team or with the people in the same project. The routine network also had the largest number of cliques. Because the routine network was also studied focusing on the most frequent ties, members in these cliques are probably tied together with relatively strong ties.

The projection of the project data onto a one-mode network creates ties between every member in the same projects. Dichotomization removes some of these ties, but the resulting one-mode network can still overestimate the level of clustering in the underlying social relations.

Work in the case company was project based, and much of the discussion of routine work should be related to the ongoing projects. The projects should then show up as similar cliques in the routine and the project networks. The results in Table 6 show that the routine and the project networks actually shared some cliques in common, but most of the cliques in the routine network were different. There can be many reasons for this. It could be that the workgroups for projects are formed of professionals of different fields, such as architects, interior designers and lighting designers. Maybe professionals of the same specialty discuss more frequently with each other across the different projects and share information with the other project participants in a different way. One reason could be the rather strict definition of cliques: all of the members must have ties with all the other members. This means that one have to talk to everyone in a project daily or more often for cliques to form. There could be more overlap in the cohesive groups in the routine and the project network, but it is not visible in lists of pure cliques.

However, the routine and the idea network did share many same cliques. That seems to suggest that how these networks formed in the first place is based on similar processes. The project network, which was derived from database entries, is formed with different logic. For example, friendship networks typically show processes such as transitivity, homophily and preferential attachment (Jackson 2010, pp.54–76). Connections between project and people arise from requirements of the projects and the business.

The self-reported networks of discussions of routine work and ideas are snapshots at a particular moment in time of participants' memory of communication during the past year. The project data, although self-reported to the database, has been constantly updated and were probably accurate on telling which projects employees worked on. Some difference could be explained by how survey respondents remembered the actual communication. Quintine and Klaumbaum (2011) calculated correlations between the self-reported and observed networks of the famous BKS studies (e.g. Bernard et al. 1970) and found correlations ranging from 0.29 to 0.43. Most of the variance was explained by other factors than who the respondent really interacted with.

The project network however is not the same as observed co-occurrences as it mostly reflects opportunities for interaction instead of true interaction. The density in the project network does not necessarily have the same interpretation as the density in the self-reported networks of actual relations, where it represents the cohesion. Actually, here it is an indication of big projects with many members. It could well mean that the bigger the project, the less frequent the interaction between members in the project because of a large number of participants and the limited time everyone has.

In directed networks, actors high in the indegree are often interpreted as popular or prestigious and actors high in the outdegree as influential (Wasserman & Faust 1994). Project participation does not have direction, either actors participate the same project or they do not. High degree in the project network does not have the interpretation that the person is popular or influential. The employees who report hours for many projects will be highly central as well as employees reporting hours to large projects with many participants. This does not tell if a person is actually central in the social structure.

6.4 Why are there observable similarities between the networks?

The routine network and the idea networks were expected to be different to each other. In many ways they were, but they also showed a high correlation between occurrences of ties. On the other hand, the routine and the project networks were expected to have structural similarities. Their similarities could be found with social network analysis, even though the similarities were not obvious at all from the graph visualizations.

As mentioned in the theoretical part, there is controversy of the network structure that best supports idea generation and creativity. Discussion of ideas involves sharing of one's knowledge on something and on the other hand asking opinion from the other party. Ideas are shared in order to get them implemented in some point of time. When actors share ideas, they expose their skills and acquired knowledge to the other party, and their ideas are evaluated. Actors are likely to share their ideas without reservation with people they trust. There are mixed results of what is the effect of psychological factors to sharing of ideas (Anderson et al. 2014). Even though creative ideas are likely to come from acquaintances outside one's near social circles (Granovetter 1973; Burt 2004; Perry-smith et al. 2006), some ideas probably require cohesion and strong ties for discussion of ideas, especially when the discussion is about to move to implementation or changes of work processes (Ohly et al. 2010).

Cohesive groups were recognized to be important for learning and sharing complex knowledge (Lave & Wenger 1991). In addition, some network members must get the new knowledge via strong ties as not everybody can be located near the structural holes. Furthermore, ideas are typically discussed together in groups, where the ideas are refined and brainstormed. This suggests that the idea network should show some similarities with other social networks, for example, some degree of clustering.

Of course, the structural holes argument or cohesive argument does not mean that ideas have to be generated only in either type of an environment. Some of the ideas occur regardless of the place or time or connections. Ideas don't necessarily have to come from the network within the organization. For example, Allen (1977) noticed that individuals who were gatekeepers by having various contact outside the organization were in an important role for the flow of new information.

Generally, high clustering in an idea network could indicate that the firm is not utilizing social networks as efficiently as it could for creativity. Both, weak ties and cohesive groups have their roles in discussion surrounding ideas. The idea network should then show sparse structure with some degree of clustering as well, but less than the routine and the project network. The results from the analysis section seem to be in alignment with this.

It could also be possible that ideas are easiest to discuss with people who are easily approached, because, for example, their similar role or position in the organizational hierarchy (Ohly et al. 2010). There is most likely already established mutual connection to these people. Establishing a connection would build trust and knowledge of other person's expertise. This may explain why there were at the same tie some similarities between the routine and the idea networks. The results in the previous section show that, although there were structural differences, there was a moderately high ($r = 0.6$) correlation between the occurrence of ties in the routine network and idea network.

The routine, idea and project network could also be similar to each other because there is another confounding factor behind. Confounding refers to external influence that affects the formation of ties and actions of the actors (Shalizi & Thomas 2011). The network studied here are special cases of communication networks in an organization, and the organization sets the context and the boundaries for the interaction and affects how the network ties can occur. The questionnaire used was designed so that the respondents first defined who they have been discussing with within the organization and then defined with whom they discussed routine work related matters and ideas.

6.5 Sources of errors and limitations

There are always sources of errors that may affect the results. Social network analysis is sensitive to non-response and missing data (Knoke & Yang 2008, pp.41–45). The response rate of the surveys was 71 %. Because the underlying relations were symmetric, the strategy used here was to assume that the missing respondent would have answered similarly to how the others respondent about him. The respondents often had different perception of frequency of communication between other members and sometimes about existence of ties. Missing ties and problems of recall could affect the results. Kossinets (2006) examined the effect of response rate to various global network properties. He found that a response-rate of 70% or better is sufficient to achieve unbiased results of in collaboration networks.

Self-reported networks might suffer from cognitive biases. For example, respondents might be risk averse and hesitate answering to questions or manipulate their answers if they have reason to suspect how the data might be used later. The questions in this survey were relatively neutral, and respondents were reminded that there are no right answers for the questions.

Self-reported ties usually differ from those observed. Bernard, Killworth and Sailer assessed reliability of informants by observing communication behavior of the groups of people and comparing the results to the self-reported interactions (e.g. Bernard et al. 1970). Their results showed that about half of what people recalled of their own interactions was in some way incorrect. The biases in recall might arise from the large amount of data that have to be remembered to answer accurately to questions of social relations (Kilduff et al. 2008). The errors made occur in two ways: some social ties are forgotten, and some ties are reported even though they did not exist (Freeman et al. 1987). Sometimes respondents tend to overstate their ties to high-status persons and understate ties to physically distant alters (Quintane & Kleinbaum 2011).

Informant inaccuracy might not be too serious a problem if the interest is on the stable patterns of repeated events. Recall of interactions in one particular event is generally poor, but Freeman, Romney and Freeman (1987) noticed that instead of showing random errors in recalling, people tended to bias towards the long-term patterns of social interaction. It seems to be easy to forget people who participate to some recurring events irregularly and falsely remember interactions with people who attended regularly, even if interaction with that person did not happen in that particular event. The question of "who did you interact with" easily gets answered by "who were you likely to interact with" (Freeman et al. 1987).

In many sections in this thesis, it was mentioned that weak ties are important for new ideas. It is not easy to define which ties in the network should be considered as strong ties and which as weak ties. Survey respondents mentioned discussing on ideas on a daily basis with many other people. Would these still be weak ties? They are definitely stronger ties than the more infrequent communication, but they might still be work-related ties, without much socializing and emotional connection. If we were interested in novelty of the ideas, maybe the focus should be on the less frequent exchanges on ideas. The available survey data did not allow to gain insight what kind of ideas are exchanged and who invented them.

Possible data processing related errors might arise from dichotomization and converting project workhour data to one-mode network. There are difficulties in choosing the right dichotomization level, and there is a lack of theory and rigorous methods for this problem (see Thomas & Blitzstein 2011b; Thomas & Blitzstein 2011a). In this study, the focus was more on the relative importance of measures than on their absolute levels. Because network densities were kept on nearly equal levels, the measures are likely to give a correct ranking for the results with the dichotomization levels used.

Creating a one-mode network of the project data brings omission errors because some interactions are left out that were not reported to workhour database. Edge attribution errors arise form assigning edges between actors because they participated the same project, but who actually did not interact with each other (Borgatti et al. 2013). This type of error cannot be avoided in networks derived from archival data. The projects in the data were only known by their ID-codes. Some projects should maybe have been removed from the data, but it was not possible without knowing the details of the projects. Holidays were, of course, removed from the data.

6.6 Theoretical Implications and future research

The thesis is based on the data from a case study and not on data gathered by random sampling of firms from some larger population. The results apply to this case organization and cannot be easily generalized to the population it represents.

The results section showed that at least in this case company, employees' co-occurrences in projects formed a network which shared more similarities with the network of discussing on routine work than network of discussing on ideas. Projecting two-mode data onto one-mode network causes clustering that might not be present in the underlying social relations. Clustering could lead to wrong inferences unless interpreted carefully.

The routine work discussion and discussion of ideas shared similarities in their network structure. Some amount of clustering in the idea network shows that cohesion does also have a role in how ideas are discussed in social networks. The many similarities of the routine and the idea networks suggest that there is an underlying social structure, which the routine network and the idea network are special cases of. Probably people choose contacts for various different reason, and having established a connection with them, evaluate if that relationship is worth keeping and what is beneficial to share in it. These different choices would form different relations with partly correlating properties but with differences as well. As this study was cross-sectional, the results showing similarities between networks does not tell anything about the causality. Seeing, for example, if working in the same project causes ties form in the discussion networks would require longitudinal data from several points in time.

Apart from the interpretation of the betweenness centrality measure (Freeman 1979), the analysis methods used in this thesis do not directly indicate structural holes or actors with brokerage opportunities. Burt (1992) has defined a set of measures for structural holes based on ego-networks. These could be used for future research to spot structural holes in the idea network and see if the data gives support for the structural wholes view to idea generation.

The dataset used in this study did not contain information on what kind of ideas were shared and who were perceived to be important for the creation of ideas. This additional data would allow to compare what kind of structural positions those actors occupy in idea network who were perceived to be creative.

The scope of this study was limited to the structural properties of the network, with the exception that information of the organizational units was used to form subsets of the data. The study could be improved by taking node attributes into the analysis. Literature of creativity shows that individual-level attributes such as cognitive capabilities, expertise, domain knowledge of the field, personality characteristics and willingness to share ideas are found to be important for people to be recognized as innovative in organization (Parzefall et al. 2008). Cultural attributes of the organizational culture are important too. Some features of the culture such as perceived competitiveness of the environment and how creativity is appreciated affect with how routine work and ideas are discussed (Perry-Smith &

Shalley 2003). Background data would allow to check, for example, is there more homophily in routine work discussion than in sharing ideas. The weak-ties view to sharing of ideas suggests that novel ideas come from actors different to oneself (Granovetter 1973; Burt 2004).

Methodologically the comparison of the three networks could be improved by studying them with exponential random graph models or ERGM models (Robins et al. 2007). The exponential random graph model allows inferences that can explain what types of social processes are important in that particular network (Lusher et al. 2012). Applied to the research question in this thesis, ERGMs could be used to model structural tendencies such as degree distribution, homophily and transitivity as endogenous variables. After controlling for these effects, it is possible to test various attributes. Attributes that could affect how ties form in idea and routine networks are for example effects of proximity, difference in tenure or organizational position.

7 Conclusion

This thesis focused on the structural characteristics of the intra-organizational social networks formed from discussion on routine work, discussion on ideas and participation in the same projects. The two discussion networks were self-reported, and the project network was derived indirectly from the time tracking database. It was found that although the project participation network was formed differently to the two self-reported networks, it shared more similarities with the network of discussing routine work than network of discussing ideas. Although the two discussion networks were subsets of an intra-organizational communication network and in that way shared similarities, the network of discussing routine work had structural differences to the network of discussing ideas.

From a managerial point of view, the study also gives a secondary result demonstrating the value of social network analysis methods. Simple descriptive statistics and visualization are often enough to give valuable insights. In fact, perception of the network outside own near social circles usually differs from how the other network members see the relationships (Krackhardt & Hanson 1993; Kilduff et al. 2008). Companies have a lot of data related to employees' work, and this can be used to reveal the social relations implied by the work design. Care must be taken in interpreting the resulting social networks as they represent of co-occurrences instead of the real social structure and show exaggerated picture of social clustering in the organization. Although this study was not longitudinal and causality cannot be seen, the similarities between the project network and the routine network suggest that how the work is organized affects the informal networks in the organization. In addition, how ideas are discussed in the organization affects the firm's ability to innovate. Analyzing the structure of the idea network can help to understand how innovations emerge. Tight clustering around the projects or strong structural similarity of routine work discussion and discussion of ideas could indicate that more diverse contacts could bring fresh ideas. In this case managers could try to create opportunities for interacting with people with various backgrounds.

8 References

Allen, T.J., 1977. *Managing the flow of technology*, Cambridge, Massachusetts: MIT Press.

Amit, R. & Belcourt, M., 1999. Human resource management processes: A value creating source of competitive advantage. *European Management Journal*, 17(2), pp.174–181.

Anderson, B.S., Butts, C. & Carley, K., 1999. The interaction of size and density with graph-level indices. *Social Networks*, 21(3), pp.239–267.

Anderson, N., Potocnik, K. & Zhou, J., 2014. Innovation and Creativity in Organizations: A State-of-the-Science Review, Prospective Commentary, and Guiding Framework. *Journal of Management*, 40(5), pp.1297–1333.

Axtell, C.M. et al., 2000. Shopfloor innovation: Facilitating the suggestion and implementation of ideas. *Journal of Occupational and Organizational Psychology*, 73(3), pp.265–285.

Barrat, A. et al., 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11), pp.3747–52.

Becker, M.C., 2004. Organizational routines: a review of the literature. *Industrial and Corporate Change*, 13(4), pp.643–678.

Bernard, H.R., Killworth, P.D. & Sailer, L., 1970. Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 80(2), pp.191–218.

Blyler, M. & Coff, R.W., 2003. Dynamic capabilities, social capital, and rent appropriation: ties that split pies. *Strategic Management Journal*, 24(7), pp.677–686.

Borg, I., Groenen, P.J.F. & Mair, P., 2013. *Applied Multidimensional Scaling*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Borgatti, S. & Cross, R., 2003. A Relational View of Information Seeking and Learning in Social Networks. *Management science*, 49(4), pp.432–445.

Borgatti, S. & Everett, M., 2000. Models of core/periphery structures. *Social networks*, 21, pp.375–395.

Borgatti, S.P., Everett, M.G. & Freeman, L.C., 2002. *Ucinet for Windows: Software for Social Network Analysis*, Harvard, MA: Analytic Technologies.

Borgatti, S.P., Everett, M.G. & Johnson, J.C., 2013. *Analyzing Social Networks*, London: SAGE Publications Ltd.

Borgatti, S.P. & Foster, P.C., 2003. The Network Paradigm in Organizational Research: A Review and Typology. *Journal of Management*, 29(6), pp.991–1013.

Brass, D. et al., 2004. Taking stock of networks and organizations: A multilevel perspective. *Academy of Management Journal*, 47(6), pp.795–817.

Buchanan, D. & Huczynski, A., 2004. *Organizational Behavior: An Introductory Text* 5th ed., Upper Saddle River, New Jersey: Financial Times/ Prentice Hall.

Burt, R., 2001. Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. S. Burt, eds. *Social Capital: Theory and Research*. Sociology and Economics: Controversy and Integration series. New York: Aldine de Gruyter, pp. 31–56.

Burt, R., 2002. The social capital of structural holes. In M. F. Guillén et al., eds. *New Directions in Economic Sociology*. New York: Russell Sage Foundation.

Burt, R.S., 1992. *Structural holes*, Cambridge, Massachusetts: Harvard University Press.

Burt, R.S., 2004. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2), pp.349–399.

Coleman, J., 1988. Social capital in the creation of human capital. *American journal of sociology*, 94, pp.S95–S120.

Cross, R. & Borgatti, S., 2004. The ties that share: Relational characteristics that facilitate information seeking. In M. H. Huysman & V. Wulf, eds. *Social capital and information technology*. Cambridge, Massachusetts: MIT Press, pp. 137–162.

Cross, R., Borgatti, S. & Parker, A., 2002. Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. *California Management Review*, 44(2), pp.25–47.

Cross, R. & Parker, A., 2004. *The Hidden Power of Social Networks: Understandting How Work Really Gets Done in Organizations*, Boston, MA: Harvard Business School Press.

Csikszentmihalyi, M., 1996. *Creativity : Flow and the Psychology of Discovery and Invention*, New York: Harper Perennial.

Faust, K. & Skvoretz, J., 2002. Comparing Networks Across Space and Time, Size and Species. *Sociological Methodology*, 32(1), pp.267–299.

Feld, S. & Carter, W.C., 1998. Foci of Activities As Changing Contexts for Friendship. In R. G. Adams & G. Allen, eds. *Placing Friendship in Context*. Cambridge, UK: Cambridge University Press, pp. 136–152.

Freeman, L., 1979. Centrality in social networks conceptual clarification. *Social networks*, 1, pp.215–239.

Freeman, L.C., Romney, A.K. & Freeman, S.C., 1987. Cognitive Structure and Informant Accuracy. *American Anthropologist*, 89(2), pp.310–325.

Granovetter, M., 1985. Economic Action and Social Structure: The Problem of Embeddedness. *American journal of sociology*, 91(3), pp.481–510.

Granovetter, M., 2005. The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives*, 19(1), pp.33–50.

Granovetter, M., 1983. The strength of weak ties: A network theory revisited. *Sociological theory*, 1, pp.201–233.

Granovetter, M.S., 1973. The strength of weak ties. *American journal of sociology*, 78(6), pp.1360–1380.

Hanifan, L.J., 1916. The Rural School Community Center. *Annals of the American Academy of Political and Social Science*, 67, pp.130–138.

Hanneman, R.A. & Riddle, M., 2005. *Introduction to social network methods*, Riverside, CA: University of California.

Hansen, M., 2001. So many ties, so little time: A task contingency perspective on corporate social capital in organizations Shaul M. Gabbay & R. T. A. J. Leenders, eds. *Research in the Sociology …*, pp.21–57.

Hansen, M., 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly*, 44(1), pp.82–111.

Hansen, M., Nohria, N. & Tierney, T., 1999. What's your strategy for managing knowledge? *Harvard Business Review*, 77(2), pp.106–117.

Heider, F., 1958. *The Psychology of Interpersonal Relations*, John Wiley & Sons.

Hodgson, G.M., 2008. The concept of routine. In M. C. Becker, ed. *Handbook of Organizational Routines*. Cheltenham, UK: Edward Elgar Pub, pp. 15–26.

Hofstede, G., 1991. *Culture and organizations: software of the mind*, New York: McGraw-Hill.

Jackson, M.O., 2010. *Social and Economic Networks*, Princeton, NJ, USA: Princeton University Press.

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

Johnson, S., 1967. Hierarchical clustering schemes. *Psychometrika*, 32(3), pp.241–254.

Kadushin, C., 2011. *Understanding Social Networks: Theories, Concepts, and Findings*, Oxford University Press, USA.

Kilduff, M. et al., 2008. Organizational network perceptions versus reality: A small world after all? *Organizational Behavior and Human Decision Processes*, 107(1), pp.15–28.

Kilduff, M. & Brass, D.J., 2010. Organizational Social Network Research: Core Ideas and Key Debates. *The Academy of Management Annals*, 4(1), pp.317–357.

Kilduff, M. & Tsai, W., 2003. *Social Networks and Organizations*, London, England: SAGE Publications Ltd.

Knoke, D. & Yang, S., 2008. *Social Network Analysis* 2nd Ed., SAGE Publications.

Koestler, A., 1989. *The Act of Creation*, London: Penguin Books.

Kossinets, G., 2006. Effects of missing data in social networks. *Social Networks*, 28(3), pp.247–268.

Krackhardt, D., 1987. QAP partialling as a test of spuriousness. *Social networks*, (9), pp.171–186.

Krackhardt, D. & Hanson, J.R., 1993. Informal Networks: The Company behind the Charts. *Harvard Business Review*, 71(4), pp.104–111.

Kramar, R.M., 1999. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, (50), pp.569–598.

Laumann, E.O., Marsden, P. V. & Prensky, D., 1983. The Boundary Specification Problem in Network Analysis. In R. S. Burt, ed. *Applied Network Analysis: a methodological introduction*. Beverly Hills: Sage Publications, pp. 18–34.

Lave, J. & Wenger, E., 1991. *Situated Learning. Legitimate peripheral participation*, Cambridge: University of Cambridge Press.

Lickel, B. et al., 2000. Varieties of Groups and the Perception of Group Entitativity. *Journal of Personality and Social Psychology*, 78(2), pp.223–256.

Lusher, D., Koskinen, J. & Robins, G. eds., 2012. *Exponential Random Graph Models for Social Networks. Theory, Methods, and Applications*, Cambridge University Press.

McPherson, M., Smith-Lovin, L. & Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), pp.415–444.

Monge, P.R. et al., 1985. The Dynamics of Organizational Proximity. *Management Science*, 31(9), pp.1129–1141.

Moreno, J., 1934. *Who Shall Survive? A New Approach to Human Interrelations*, Washnington: Nervous and Mental Disease Poublishing Co.

Munroe, P.T., 2007. Cognitive Balance Theory (Heider). In G. Ritzer, ed. *Blackwell Encyclopedia of Sociology Online*. Oxford, UK, Malden, USA and Carlton, Australia: Blackwell Publishing Ltd.

Nahapiet, J. & Ghoshal, S., 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, 23(2), pp.242–266.

Nelson, R.R. & Winter, S.G., 1982. *An Evolutionary Theory of Economic Change*, Cambridge, Massachusetts: Belknap.

Newman, M.E.J., 2010. *Networks: An Introduction*, Oxford University Press.

Nonaka, I. & Tackeuchi, H., 1995. *The knowledge creating company: How Japanese companies create the dynamics of innovation.*, New York: Oxford University Press.

Obstfeld, D., 2005. Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative science quarterly*, 50(1), pp.100–130.

Oh, H. & Kilduff, M., 2008. The ripple effect of personality on social structure: self-monitoring origins of network brokerage. *The Journal of applied psychology*, 93(5), pp.1155–64.

Ohly, S., Kase, R. & Skerlavaj, M., 2010. Networks for generating and for validating ideas: The social side of creativity. *Innovation: managemet, policy & practice*, 12(1), pp.100–130.

Padgett, J. & Ansell, C., 1993. Robust Action and the Rise of the Medici, 1400-1434. *American journal of sociology*, (6), pp.1259–1319.

Parzefall, M., Seeck, H. & Leppänen, A., 2008. Employee innovativeness in organizations: a review of the antecedents. *Finnish Journal of Business Economics*, 2(8), pp.165–182.

Perry-smith, A.J.E. et al., 2006. Social Yet Creative : The Role of Social Relationships in Facilitating Individual Creativity. *The Academy of Management Journal*, 49(1), pp.85–101.

Perry-Smith, J.E. & Shalley, C.E., 2003. The Social Side of Creativity: A Static and Dynamic Social Network Perspective. *The Academy of Management Review*, 28(1), p.89.

Powell, W., 1990. Neither market nor hierarchy. In B. M. Staw & L. L. Cummints, eds. *Research in Organizational Behavior*. Sage Publications, Inc, pp. 295–336.

Prell, C., 2012. *Social Network Analysis: History, Theory and Methodology*, London: SAGE Publications Ltd.

Putnam, R.D., 1995. Bowling Alone: America's Declining Social Capital. *Journal of Democracy*, 6(1), pp.65–78.

Putnam, R.D., 2000. *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster.

Quintane, E. et al., 2013. Short- and long-term stability in organizational networks: Temporal structures of project teams. *Social Networks*, 35(4), pp.528–540.

Quintane, E. & Kleinbaum, A., 2011. Matter over mind? E-mail data and the measurement of social networks. *Connections*, 31, pp.22–46.

Robins, G. et al., 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2), pp.173–191.

Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika*, 31(4), pp.581–603.

Scott, J., 2013. *Social network analysis* 3rd Editio., London: SAGE Publications Ltd.

Shalizi, C.R. & Thomas, A.C., 2011. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological methods & research*, 40(2), pp.211–239.

Smedlund, A., 2010. Characteristics of routine, development and idea networks in teams. *Team Performance Management*, 16, pp.95–117.

Smedlund, A., 2008. The knowledge system of a firm: social capital for explicit, tacit and potential knowledge. *Journal of Knowledge Management*, 12(1), pp.63–77.

Smedlund, A. & Choi, E., 2009. Communication networks in routine and non-routine tasks. *Helsinki University of Technology, Laboratory of Work Psychology and Leadership, Working Paper no. 2009/2.*

Spence, I. & Ogilvie, J.C., 1973. A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivariate Behavioral Research*, 8(4), pp.511–517.

Spender, J.-C., 1996. Making knowledge the basis of a dynamic theory of the firm. *Strategic Management Journal*, 17(S2), pp.45–62.

Stevenson, W.B., 1990. Formal structure and networks of interaction within organizations. *Social Science Research*, 19(2), pp.113–131.

Teece, D.J., Pisano, G. & Shuen, A., 1997. Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), pp.509–533.

Thomas, A. & Blitzstein, J., 2011a. Valued Ties Tell Fewer Lies, II: Why Not To Dichotomize Network Edges With Bounded Outdegrees. *arXiv preprint arXiv:1101.2228*, pp.1–15.

Thomas, A. & Blitzstein, J., 2011b. Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds. *arXiv preprint arXiv:1101.0788*, pp.1–36.

Uzzi, B., 1997. Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, 42(1), p.35.

Uzzi, B., 1996. The Sources and Consequences of Embeddedness for theEconomic Performance of Organizations: The Network Effect. *American sociological review*, 61(4), pp.674–698.

Wasko, M.M. & Faraj, S., 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, 29(1), pp.35–57.

Wasserman, S. & Faust, K., 1994. *Social Network Analysis: Methods and Applications*, Cambridge, Massachusetts: Cambridge University Press.

Watts, D.J., 2004. *Small Worlds: The dynamics of networks between order and randomness* 8th ed., Princeton, New Jersey: Princeton University Press.

Watts, D.J. & Strogatz, S.H., 1998. Collective dynamics of small-world networks. *Nature*, 393(6684), pp.440–442.

Weick, K.E., 1990. The Vulnerable System: An Analysis of the Tenerife Air Disaster. *Journal of Management*, 16(3), pp.571–593.

Wenger, E., 2000. Communities of Practice and Social Learning Systems. *Organization*, 7(2), pp.225–246.

Appendix A.

Network Survey

The questionnaire was sent to all employees in the firm's main office. The survey questions are shown below.

Defining the sample

Question 1: Who have you been discussing with (face-to-face, email, phone etc.) with a frequency of at least once per quarter of a year?

Below is a list of persons. Depending on your work task or tenure you might know some people really well, but some people hardly at all.

Please tick those persons that you have been personally discussing with in a frequency of at least once per quarter of a year. The discussions can be either work-related or informal – coffee breaks count too.

In case you later notice that you wish to add somebody, you can come back to this page anytime during the survey by clicking "back" and "forward" buttons at the end of each page.

Ideas

This question is to find out those individuals that you exchange knowledge with related to new ideas and new identified possibilities. Below are listed characteristics that describe ideas:

- You get a "light bulb moment" and come up with an idea that is somehow related to your own work
- Idea is new in your opinion and you are not aware if anyone else has exactly similar ideas.
- Idea can be also a development proposition that relates for example to the development of working manners or business in your company.
- Everyone of us may have our own places and ways to come up with ideas. Ideas can be born in peculiar situations at work, at home or during free time.
- Ideas can be born and shared in informal situations, for example during coffee breaks

Question: Please estimate for each person that you mentioned in question 2, how often do you discuss with them about your own ideas?

Not at all	*Every day*	*Once per week or more*	*Less than once per week*

Routine work

This question is to map those individuals that you exchange knowledge with related to your routine work tasks.

The characteristics of a routine work task are for example the following:

- It can be a work task that is delivered to the client that is pre-defined and that, according to your own opinion, is a part of the routine of your competence and daily basic work
- can be a pre-defined, recurring work task that has to be done in a given timeframe
- It can be a part of the internal affairs of your company, i.e. registration of work hours or travel bills

Question: Please estimate for each person that you mentioned in question 2, how often do you discuss with them on subjects that are related to your routine work tasks?

Not at all	*Every day*	*Once per week or more*	*Less than once per week*

Appendix B.

Formula for eigenvector centrality

To derive a formula for Eigenvector centrality, Newman (2010) uses the following approach. Each tie's weight depends on the centrality measure of the node it is incident to. First, one can make an initial guess for the centrality of a node, say $x_i = 1$. Centralities can then be calculated again taking into account all other nodes $x_i{}' = \sum_j A_{ij}x_{ij}$ where $\mathbf{A}$ is the adjacency matrix. This can be written in a vector form for whole network as $\mathbf{x}' = \mathbf{Ax}$. After t steps the estimate for the centrality has changed to $x'(t) = \mathbf{A}^t\mathbf{x}(0)$. Here $\mathbf{x}(0)$ can be written with its eigenvectors $\mathbf{x}(0) = \sum_i c_i\mathbf{v}_i$, where c_i is some constant. After t steps centrality becomes $x(t) = A^t \sum_i c_i\mathbf{v}_i = \sum_i c_i\lambda_i^t\mathbf{v}_i = \lambda_1^t \sum_i c_i \left[\frac{\lambda_i}{\lambda_1}\right]\mathbf{v}_i$, where λ_1 is the largest eigenvalue. Because the fraction $\frac{\lambda_i}{\lambda_1} < 1$ when $i \neq 1$, all terms except first one diminish exponentially as $t \to \infty$ and the sum approaches to $\mathbf{x}(t) \to \lambda_1^t c_1\mathbf{v}_i$, it is proportional only to the leading eigenvalue. This can equally be written in a famous eigenvector form:

$$\mathbf{Ax} = \lambda_1\mathbf{x}$$

.

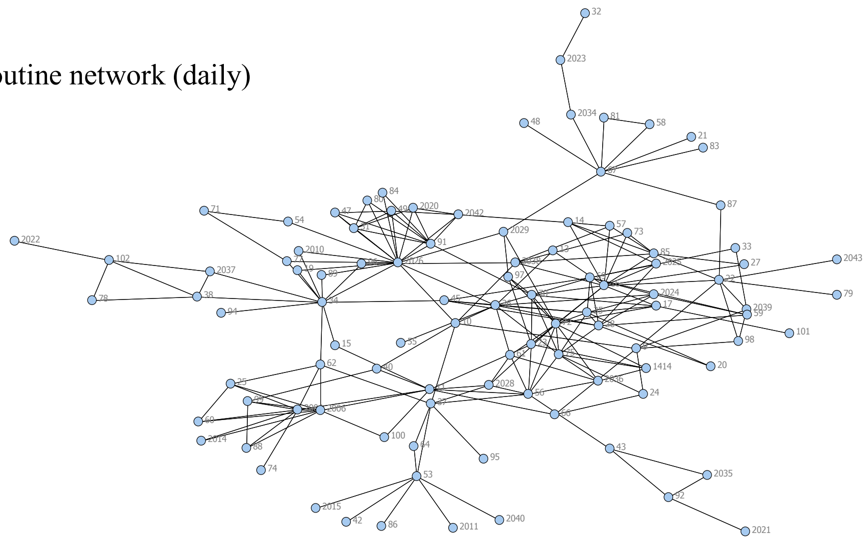
Eigenvector centrality works well for simple graphs as their adjacency matrices are symmetric. For asymmetric matrices formulas must be modified to take into account the two set of eigenvalues. Another measures for node importance are Katz centrality and Google PageRank. Both of them work similarly to eigenvector centrality with some adjustments related to directional networks. (Newman, 2010)

The four different centrality measures discussed here can be summarized by their different point of view they take to centrality. Degree centrality can be seen as measuring activity of an actor, Eigenvector centrality measures the extend an actor is surrounded by highly connected actors, betweenness measures potential control for information flow and closeness centrality indicates actor's independence of information channels. (Prell 2012, p.108)

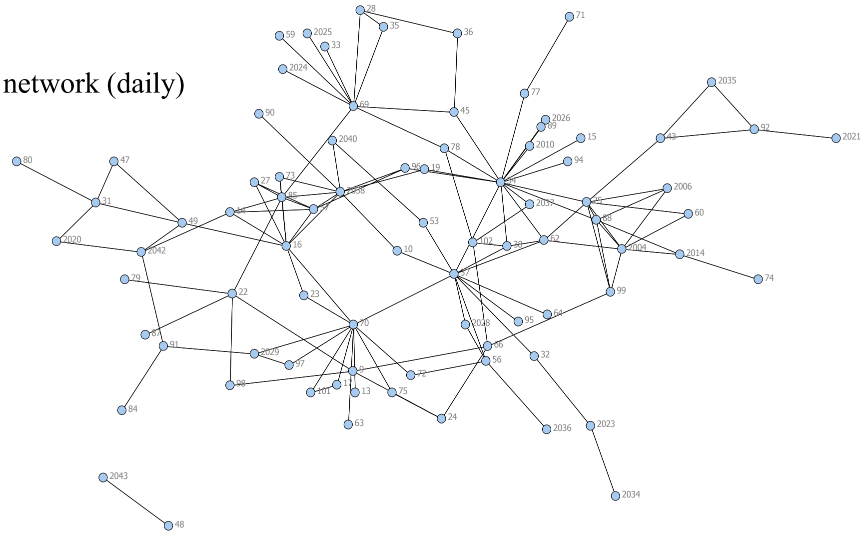
Appendix C.

Graph visualizations of the routine, idea and project networks

Routine network (daily)

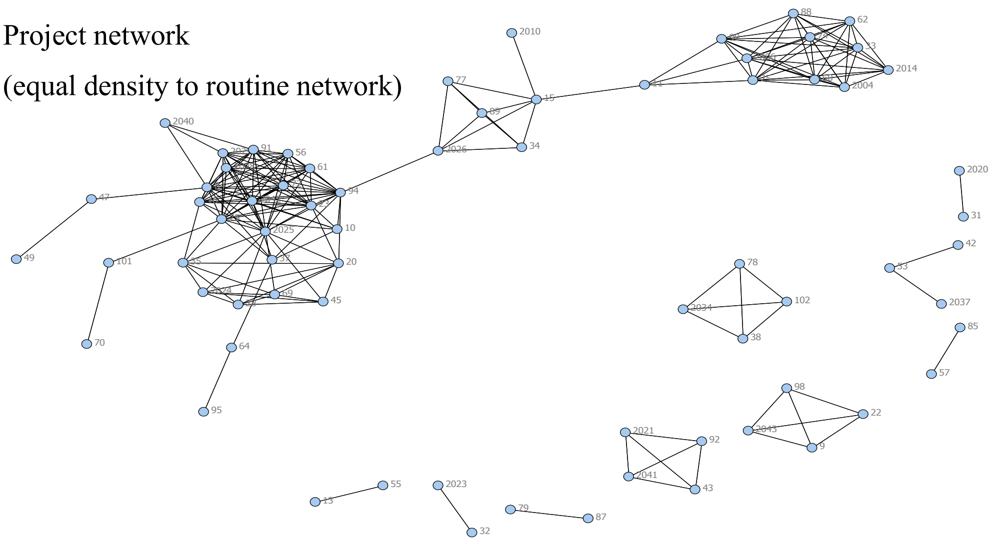


Idea network (daily)



Project network

(equal density to routine network)



Appendix D.

Additional tables from analysis

Within and between unit densities in each network at daily communication frequency. Project has been dichotomized to have same density as routine network.

Discussion on routines (daily)

UNIT	1	2	3	5	6	7	8	9	10
1	0.121	0.009	0.021	0.007	0.006	0.011	0.054	0.005	0.000
2		0.500	0.028	0.037	0.034	0.007	0.000	0.012	0.020
3			0.500	0.063	0.000	0.017	0.036	0.042	0.000
5				0.152	0.013	0.033	0.030	0.046	0.000
6					0.103	0.015	0.000	0.051	0.014
7						0.105	0.024	0.048	0.018
8							0.198	0.016	0.006
9								0.124	0.025
10									0.091

Discussion on ideas (daily)

UNIT	1	2	3	5	6	7	8	9	10
1	0.000	0.019	0.042	0.000	0.006	0.000	0.042	0.009	0.008
2		0.167	0.000	0.028	0.000	0.000	0.008	0.006	0.010
3			0.667	0.042	0.019	0.017	0.018	0.014	0.000
5				0.061	0.019	0.000	0.018	0.005	0.000
6					0.090	0.015	0.000	0.051	0.014
7						0.067	0.024	0.019	0.012
8							0.176	0.016	0.006
9								0.085	0.015
10									0.000

Project participation (daily)

UNIT	1	2	3	5	6	7	8	9	10
1	0.045	0.009	0.000	0.035	0.038	0.033	0.060	0.056	0.098
2		0.000	0.028	0.019	0.009	0.015	0.000	0.012	0.040
3			0.000	0.042	0.000	0.017	0.000	0.014	0.000
5				0.015	0.038	0.028	0.036	0.032	0.038
6					0.038	0.026	0.038	0.043	0.042
7						0.019	0.038	0.033	0.061
8							0.022	0.036	0.058
9								0.059	0.045
10									0.109