**University of Zürich**

Department of Business Administration  
Institute for Network Science  
Prof. Dr. Claudio J. Tessone

Master thesis

**Mapping information spreading to   
improve communication:   
The case of ONE smart solution**

[Beschreibender Untertitel]

**Made by:**

Denis Krebs  
Herzogstrasse 9  
5000 Aarau  
077 411 00 14  
denis.krebs@uzh.ch  
Student ID: 11-747-912  
Faculty of Business, Economics and Informatics  
Number of Semesters: 13

Supervisor: Prof. Dr. Claudio J. Tessone  
Advisor: Dr. Zhao Yang

Submitted on: 29/12/2018

* 4 Bewertung einer wissenschaftlichen Arbeit

Bei der Bewertung der Arbeiten werden die folgenden 3 Bereiche geprüft:

1) Formaler Aufbau

Beurteilt werden Zitierweise, Abbildungen, Literaturverzeichnis, Formatierungen und Orthographie/

Interpunktion die Vorgaben der Wegleitung.

2) Struktur

Ist ein roter Faden in der Arbeit erkennbar und wurde systematisch auf die Beantwortung der

Fragestellungen hingearbeitet? Sind die einzelnen Kapitel und Unterkapitel ausgewogen?

3) Inhalt

Wurde das formulierte Ziel der Arbeit erreicht und die daraus abgeleiteten Fragestellungen

beantwortet? Wurde der aktuelle Forschungsstand anhand wissenschaftlicher Beiträge aufgezeigt?

Sind die Gedankengänge klar und die Wortwahl einer wissenschaftlichen Arbeit entsprechend?

Wurde die relevante Literatur zum Thema der Arbeit verarbeitet?

TIPP: Eine detaillierte Checkliste für wissenschaftliche Arbeiten, welche die 3 genannten

Bewertungsbereiche noch konkretisiert und mit deren Hilfe man die eigene Leistung genau

überprüfen kann, findet sich auf der nächsten Seite (Abbildung 8).

**Abstract**

…

Table of contents

[Table of figures VI](#_Toc533782528)

[List of abbreviations VII](#_Toc533782529)

[1 Introduction & Problem Statement - 1 -](#_Toc533782530)

[1.1 Scientific Research Statement - 1 -](#_Toc533782531)

[1.2 Research context and relevance - 1 -](#_Toc533782532)

[1.2.1 Scientific context - 1 -](#_Toc533782533)

[1.2.2 Business relevance - 2 -](#_Toc533782534)

[1.3 Terminology / structure of the thesis - 3 -](#_Toc533782535)

[2 Theory - 5 -](#_Toc533782536)

[2.1 Network Science - 5 -](#_Toc533782537)

[2.1.1 Characteristics of Network Science - 6 -](#_Toc533782538)

[2.1.2 Networks at the Heart of Complex Systems - 8 -](#_Toc533782539)

[2.1.3 Two Forces Helped the Emergence of Network Science - 8 -](#_Toc533782540)

[2.2 Tools - 9 -](#_Toc533782541)

[3 Research State & gaps - 15 -](#_Toc533782542)

[3.1 Email network analysis - 15 -](#_Toc533782543)

[3.2 Connectivity patterns - 15 -](#_Toc533782544)

[4 M-Industry & Project ONE - 18 -](#_Toc533782545)

[4.1 Migros Group - 18 -](#_Toc533782546)

[4.2 Migros Industry - 18 -](#_Toc533782547)

[4.3 ONE - 18 -](#_Toc533782548)

[5 Hypothesises - 20 -](#_Toc533782549)

[6 Concept & Method - 21 -](#_Toc533782550)

[6.1 Data - 21 -](#_Toc533782551)

[6.1.1 Email log - 21 -](#_Toc533782552)

[6.1.2 Questionnaire - 22 -](#_Toc533782553)

[6.2 Operationalization of the research question - 22 -](#_Toc533782554)

[6.3 Execution - 23 -](#_Toc533782555)

[6.3.1 Static network - 24 -](#_Toc533782556)

[6.3.2 Dynamic network - 25 -](#_Toc533782557)

[6.3.3 In-block nestedness - 26 -](#_Toc533782558)

[7 Results - 27 -](#_Toc533782559)

[7.1 Description of population & sample - 27 -](#_Toc533782560)

[7.1.1 Questionnaire - 27 -](#_Toc533782561)

[7.1.2 Social networks - 28 -](#_Toc533782562)

[7.2 Research questions & answers - 30 -](#_Toc533782563)

[How does SNA help to map and understand the flow of communication? - 30 -](#_Toc533782564)

[H1 & H2: A lot of communication within subproject and weak communication between subprojects - 33 -](#_Toc533782565)

[H3: Lack of information for SG about other subprojects - 35 -](#_Toc533782566)

[How does SNA help to understand the connections between individuals? - 35 -](#_Toc533782567)

[7.3 Summary of results - 36 -](#_Toc533782568)

[8 Conclusion - 37 -](#_Toc533782569)

[8 Recommendations for ONE - 38 -](#_Toc533782570)

[Insights to help to improve communication - 38 -](#_Toc533782571)

[Bibliography - 39 -](#_Toc533782572)

[Appendix - 47 -](#_Toc533782573)

[Achtung: Manuell Zeilenumbruch zwischen letztem Kapitel und Literaturverzeichnis einfügen]

# Table of figures

[Figure 1 Bridges of Konigsberg. Four landmasses connected by six bridges. - 6 -](#_Toc533856163)

[Figure2: Visualization of an Example Social Network - 11 -](#_Toc533856164)

[Figure 3: [...] - 14 -](#_Toc533856165)

[Figure 4 Mails per day during the period under consideration. The many sharp drops are the weekends on which less mails are sent. - 28 -](#_Toc533856166)

[Figure 5 Slicing the ONE dataset. Each event is an email communication within the ONE project team. We can see that internal length reacts to the rate of change of events. The similarity reflects the rate of change of events at other times - 29 -](#_Toc533856167)

[Figure 6 Dynamic network with DynSnap timescales. Number of vertices are raising constantly. Edges represent a connection between two individuals (number of vertices in undirected and unweighted network). Number of mails equals the number of edges in directed and weighted network). - 29 -](#_Toc533856168)

[Figure 7 Complete ONE network (simplified). Including 595 vertices and 19'037 edges. Lines represent an interaction (at least one email) between two actors. Vertices placed on the plane using the force-directed layout algorithm by Reingold & Fruchterman (1991). - 30 -](#_Toc533856169)

[Figure 8 Node level SNA measures on the DynSnap timescales. Top row shows the evolution of assortativity in simplified (left) and un-simplified (right) network. Middle shows density (left) and transitivity (right). Bottom row is about path length, dimeter (left) and the average path length (right). - 32 -](#_Toc533856170)

[Figure 9 Centrality measures. Distribution of nodes. Degree centrality (left) and closeness centrality (right). - 33 -](file:///C:\Users\Denis%20Krebs\Desktop\Thesis\Thesis_Working%20Document_20181228.docx#_Toc533856171)

[Figure 10 Centrality measures on DynSnap networks. Freeman's measures of centralization: degree (left), betweenness (middle) and closeness (right). expression of how tightly the graph is organized around its most central point. Three types of graph centralization to capture the various aspects of centralization. - 33 -](file:///C:\Users\Denis%20Krebs\Desktop\Thesis\Thesis_Working%20Document_20181228.docx#_Toc533856172)

[Figure 11 Mutual information measure of Louvain communities compared to company of origin (left) and manually assigned subproject (right) for site group members. - 34 -](file:///C:\Users\Denis%20Krebs\Desktop\Thesis\Thesis_Working%20Document_20181228.docx#_Toc533856173)

[Figure 12 Mutual information measure of Louvain communities compared to company of origin (left) and assigned subproject (right) for core team members. - 34 -](file:///C:\Users\Denis%20Krebs\Desktop\Thesis\Thesis_Working%20Document_20181228.docx#_Toc533856174)

[Figure 13 In-block nestedness for the DynSnap timescales for the ONE network, Erdös-Renyi model and Configuration model. Values for the ONE dataset are significantly different from the two random model indication in-block nestedness. - 35 -](file:///C:\Users\Denis%20Krebs\Desktop\Thesis\Thesis_Working%20Document_20181228.docx#_Toc533856175)

[automatisches Abbildungsverzeichnis, rechte Maustaste 🡪 Felder aktualisieren]

[neue Abbildung im Text einfügen: Verweise 🡪 Beschriftung einfügen]

List of abbreviations

SNA Social Network Analysis

MGB (FME) Migros-Genossenschafts-Bund (Federation of Migros Cooperatives)

M-Industry Migros Industry

ONE One smart solution

ERP Enterprise Resource Planning

PMO Project Management Office

IBN In-block nestedness

NMI Normalized mutual information

[Tipp: Bei der ersten Verwendung der Abkürzung im Text diese in Klammern ausschreiben und kurz erklären]

# 1 Introduction & Problem Statement

The success of companies greatly depends on understanding the informal web of contacts because these hidden connections drive how work gets done. Already 2005 Kate Ehrlich and Inga Carboni state, that “A social network analysis examines the structure of social relationships in a group to uncover the informal connections between people.” and therefore is an effective method for revealing the hidden connections between people. (Ehrlich & Carboni, 2005)

## **1.1 Scientific Research Statement**

In this thesis social network analysis is applied on an empirical dataset to unveil hidden patterns of human interaction. We conduct basic SNA measures, community detection with the Louvain algorithm as well as the relatively recent approach of in-block nestedness from Solé-Ribalta, Tessone, Mariani, & Borge-Holthoefer (2018).

With the analysis of a real-world dataset, beside answering the following research questions, theoretical models can be applied, proved and eventually extended (not the primary goal of the master thesis).

1. Can network analysis help to map and understand the flow of communication in a project in a real business environment?
2. How does network analysis help to understand the connection between individuals working in a big business project?
3. With which insights about how information in an organization spread – generated with network analysis out of the data from the project ONE smart solution – can the communication within the core project team be improved? Whereas improved means: The collaborator know where they can get the relevant information (for themselves) about the subprojects of ONE smart solution.

## 1.2 Research context and relevance

How does the thesis contribute to bring the research state a step ahead and how it distinguishes itself from already existing literature?

### 1.2.1 Scientific context

Since network analysis is a relatively new research field (Newman, Barabási, & Watts, The structure and dynamics of networks, 2006, p. 4), with an increasing relevance since the beginning of the 21st century and especially in recent years (National Research Council (U.S.), 2005, p. 11). The rise of the internet, which basically is a large network, and online social networks lets suggest that this boom won’t be over soon (Contreras Arias, 2017, p. 39). Not only the internet is a network or contains them (social media networks), but networks are everywhere (Barabási & Pósfai, 2016a, p. 1). There is an increasing number of theoretical studies in the field of network science, which mostly address scientific problems in biology (genetic networks), artificial intelligence (neural networks), geography, transport networks and sociology (diseases spreading) (Otte & Rousseau, 2002, p. 441).

The work addressing social networks is mostly focusing either on co-authorship networks in the scientific world such as the work of Cheong & Corbitt (2009) or on social media network analysis (Himelboim, 2017), but not the communication network within a particular business or project (Newman, Barabási, & Watts, The structure and dynamics of networks, 2006, p. 4). An interesting real-world application of SNA are emerging patterns of collaboration, because a better quality of communication within organizations is linked to higher levels of performance and service (Quelle: Auditing organizational communication : a handbook of research, theory and practice). Nevertheless, social networks included in email logs thought to date have been little explored and many studies testing the approach of social network science in the organizational context analyse the Enron dataset (Tran & Khaw, 2006).

The contribution of this thesis should exactly be here; we apply existing methods to a new real-world communication dataset and providing further evidence.

### 1.2.2 Business relevance

The master thesis shows how information in business flows and maps the actual communication with network analysis. The question “Who communicates with who” occupies every project manager (Welch & Jackson, 2007) and I want to answer it for M-Industry’s project ONE smart solution. An analysis like this was never made inside the “Migros universe” before. Formal relations represented in organizational charts are important and structure certain kinds of communications, informal relationships and employee interaction defines how work is accomplished (Borgatti & Molina, n.d., p. 1). In the past, companies interested in tracking the flow of information might conducted a survey and interviewed employees. These methods are inferior, since they are biased and people only reveal the information they want to share (Quelle: Problem mit Interview als Datenquelle). Reliable data should be systematically collected by the system, regardless of the study conducted, so the behaviour of the examinee is unbiased. This is the case with emails. Every mail is stored in the servers of the IT department of Migros (MITS) and we worked hard to get access to those real and unbiased treasure of information. the thesis will provide information to the strategically important project and helps that ONE can be success-fully implemented into the business.

The master thesis helps to tackle a problem al-most every big project has – track how information flows and find the best communication path to spread information – in a new way.

## 1.3 Terminology / structure of the thesis

According to Newman (2001) a social network is:

“a collection of people, each of whom is acquainted with some subset of the others. Such a network can be represented as a set of points (or vertices) denoting people, joined in pairs by lines (or edged) denoting acquaintance.”

Social networks can represent any community no matter what size it has, as for example a company or a firm, a school or the entire world. (Newman, 2001, p. 1)  
In this thesis a graph related representation is used to systematically describe the relations between people (points). Graphs were existent long before social network analysis became a thing and originate from mathematics and computer science. Graphs can represent all kind of networks so that SNA directly benefits from the theoretical work in this two research fields.

The interacting elements of a social network (individuals), the fundamental units of which graphs are formed, are called nodes, vertices or actors and are represented with points. The second fundamental unit are the connections between objects, called edges. The connections, usually lines between nodes, can represent friendship, mail communication or other types of interactions. Vertices connected by an edge are called neighbouring or adjacent. (Mariani, 2017, pp. 9, 10) Edges can be directed (e.g. in online social networks you can follow a celebrity but this doesn’t necessarily mean that the celebrity also follows you) or undirected (friendship between two actors normally is reciprocal). The interaction intensity between two vertices is called weight (e.g. number of mails sent between two people) (Kunc, n.d., p. 220).

Graphs appear in several varieties which are defined by the type of relationships they represent. In this text the words ‘network’ and ‘graph’ are synonymous even if network contains additional information about the vertices compared a graph. Such additional information as age, sex, etc. can be display in a graph as well (e.g. colouring the nodes according to the age).

The rest of the thesis is structured as followed: An introduction to the theory of network science, the tools to model social networks, the scientific research state and gaps make the beginning of the main part of the thesis. After maintain a basic understanding about the science part, the practical environment with the Migros Group, M-Industry and the project ONE will be outlined so that the business context becomes clear. The further parts are about the analysis conducted, starting with the hypothesises and the underlying data. The method chapter describes the data and the operationalization of the research question, which is split into two parts, the quantitative mail part and the qualitative survey. These parts are brought together in the results section where we answer the research questions stated before. The conclusion connects the theory and the conducted analysis. In the last part we bridge them to the real world by elaborating recommendations for ONE smart solution.

# 2 Theory

Network Science is the base for some of the most successful companies of the 21st century. Facebook, Twitter, LinkedIn, Cisco, Apple and Google base their technology on networks. The business models either make networks visible (Facebook, Twitter, LinkedIn) or use them to facilitate fast and simple access to desired information (Google).

Attention towards networks and in particular social network analysis experienced a remarkable growth in recent years (Garton, Haythornthwaite, & Wellman, 2006). Networks are applied in many different research disciplines such as computer science (World Wide Web) , artificial intelligence (neural networks), geography and transport networks, economics and disease transmission (Barabási, Newman, Watts, Barabasi, & Watts, 2006, p. 1). Mathematical graph theory is the underlaying basis of the multidisciplinary approach (Otte & Rousseau, 2002, p. 2). Graphs are a structure studied by mathematicians since 1736 when Euler solved the problem of the Königsberg bridges (Barabási et al., 2006, p. 1). SNA answers social research questions based on precise formal mathematical definitions (Kirchhoff, 2010) as for example co-authorship networks, collaboration structures and other forms of social interactions (Otte & Rousseau, 2002, S. 24).

Social network analysis (SNA) focuses on the structure of ties within a set of social actors, e.g., persons, groups, organizations, and nations, or the products of human activity or cognition such as web sites. The guiding assumption of SNA is that the way used by the members of a group to communicate to each other affects some important characteristics of that group, e.g. efficiency when performing a task, moral satisfaction, leadership (Knoke, 2008; Scott, 2012).

## 2.1 Network Science

Network science has its origin in graph theory which dates back to the 18th century. 1736 Leonard Euler published the first paper in history of graph theory – “Geometry is unimportant, only degree matters”. Euler tried to solve the puzzle of the bridges of Königsberg (today Kaliningrad). The city of Königsberg consists of four landmasses and seven bridges. He tried to answer the question “is there a trail that transverses each bridge exactly once?”, by using graph theory. Landmasses were represented as nodes and the bridges as edges. (Mariani, 2017, p. 5)

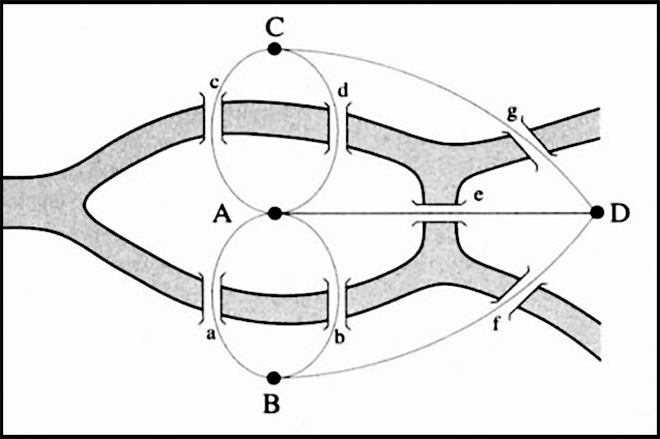


Figure 1 Bridges of Konigsberg. Four landmasses connected by six bridges.

Network analysis is a relatively new research field (Newman, Barabási, & Watts, The structure and dynamics of networks, 2006, p. 4), with an increasing relevance since the beginning of the 21st century and especially in recent years (National Research Council (U.S.), 2005, p. 11). The rise of the internet, which basically is a large network, and online social networks lets suggest that this boom won’t be over soon (Contreras Arias, 2017, p. 39). Not only the internet is a network or contains them (social media networks), but networks are everywhere (Barabási & Pósfai, 2016a, p. 1). There is an increasing number of theoretical studies in the field of network science, which mostly address scientific problems in biology (genetic networks), artificial intelligence (neural networks), geography, transport networks and sociology (diseases spreading) (Otte & Rousseau, 2002, p. 441). Biologists used network theory to examine the metabolic activities of cells in different micro-organic species (Jeong, Tombor, Albert, Oltvai, & Barabási, 2000).

### 2.1.1 Characteristics of Network Science

Network: - an interconnected group or system: a network of shops   
- Also: net a system of intersecting lines, roads, veins, etc   
- computing a system of interconnected computer systems, terminals, and other equipment allowing information to be exchanged   
(2018 Dictionary.com, LLC)

The study objects – a network – but also the methodology define the approach of network science. The following four aspects define the nature of network science methodology. In combination it offers multi-faceted tools and perspectives necessary to understand the properties of real networks. (Barabási & Pósfai, 2016b)

**Interdisciplinary**

Network science is not limited to one discipline, rather offers a language allowing different disciplines to interact. The common nature of many issues cell biologists, brain scientists and computer scientist deal with has led to a cross disciplinary fertilization of tools and ideals. Each trying to characterize the wiring diagram behind their system, too extract information of a noisy and incomplete dataset or trying to understand the robustness to failures.

**Empirical, Data Driven**

Graph theory, a field of mathematics, is the root of many key concepts in network science. Unlike graph theory, network science is empirical and is never satisfied with abstract mathematical tools to describe network property. It focuses on testing the tools on real data. Insights about the network’s properties and behaviour determine utility.

**Quantitative and Mathematical**

One of the key concepts borrowed from graph theory is the formalism to deal with graphs and to properly use its tools. The conceptual framework to seek universal organizing principles is and the way to deal with randomness is derived from statistical physics.  
Theoretical formalism must be understood to efficiently master the tools, even if network analysis software has opened their use up to a broad community.

**Computational**

Many or most practical networks are big in size: Social networks include every human being and the neural networks consists of billions of neurons. The shear amount of auxiliary data regularly confronts network scientists to computational challenges. Algorithms, database management and data mining serve to tackle the strong computational character.

### 2.1.2 Networks at the Heart of Complex Systems

The world around us in the 21st century consists of complex systems. Global society is the cooperation between billions of individuals, communication infrastructure is integrated from cell phones and computers which are interconnected and even our brain requires the coherent activity of billions of neurons to let us comprehend our world. (Barabási & Pósfai, 2016a) These and many other cotemporally systems are called complex systems, whereas complex describes something that consists of many interconnected parts and is characterized by a very complicated or involved arrangements of parts (Dictionary.com, 22/11/2018). The emergence of network science is an approach to develop a deep understanding of complex systems, mathematically describe, predict and eventually control them.   
Behind all complex systems is a network in which the interactions between the components are hidden. The underlaying networks are driven by a common set of fundamental laws and organizing principles, disregarding the diversity of complex systems. Once we abstract away the exact nature of interactions and components, networks, no matter the size, nature, age or scope, are more similar than different. (Barabási & Pósfai, 2016a)

### 2.1.3 Two Forces Helped the Emergence of Network Science

Networks themselves are not new. Metabolic networks have a history of billions of years dating back to the origins of life, social networks are as old as humanity and mathematicians exploring graphs since 1736.

As stated above a detailed network map is necessary to describe the behaviour of a complex system. In the case of a social system this requires a list of your friends, your friend’s friends, and so on. In biology the map must contain binding interactions and chemical reactions involving genes and proteins. Before the internet revolution we lacked the availability of data and the tools to map and track the huge amount of data behind complex systems. The technological advances allowed to create diverse maps from complex systems (e.g. Facebook and LinkedIn developed accurate depositories of friendships or professional ties). These diverse network maps allowed network scientists to overcome the differences of various networks we encounter in nature and science (e.g. nodes can be molecules in a metabolic network or individuals in social systems). A key discovery is the universal properties behind the different systems. The architecture of networks, disregarding the domain they’re emerging of, is similar to each other. This universality offers the foundation of the new discipline of network science. (Barabási, 2014, pp. 2–3)

## 2.2 Tools

A graph is a mathematical object. Formally, Graph theory defines a network as the mathematical notation , where V is the vertex set (actors) and E is the edge set (relations or interactions) of the community. Networks are representations of real-world (complex) systems, where the context adds a specific meaning to nodes and links. Graph techniques can be used to analyse real-world networks, whereas all networks are graphs but not the other way around. (Mariani, 2017, p. 13)

Social networks describe a group of people, in which each individual is acquainted with a subset of the others. The wiring structure of a network (who is connected to who) has important implications on the behaviour. For example, a variation in the average number of acquaintances that individuals have, the average degree of the network, influences the spread of information. By abstracting away details, graph theory helps social scientists to understand empirical data and describe their structure. A social network is represented as a set of points connected by lines. The points represent people which are connected by a line if they interact with each other (Newman, Barabási, & Watts, The structure and dynamics of networks, 2006, S. 221ff). The number of nodes is called the size of the network (N). Number of links (L) represents the total number of interactions between all nodes. High-degree vertices are often called hubs. In social network science we use the terms network, node, link which are interchangeably with the terms graph, vertex and edge from graph theory. (Barabási & Pósfai, 2016)

In social network analysis (SNA) seeks to predict relationships shaping the interaction between entities with quantitative analysis (Butts, 2008, p. 1). The specific nature of the entities depend on the object of study and can be persons (our case), groups or organizations (Butts, 2008, p. 2). SNA has four underlaying key assumptions, defined by Freeman (2004): i) motivated by “structural intuition” based on the linking ties between actors; ii) based on empirical data which is systematically collected; iii) based on mathematics and/or computational models; iv) largely relies on graphics. By quantifying social structures SNA can identify important actors, therefor it is also referred to as structural analysis (Wellman & Berkowitz, 1988). SNA answers social research questions based on precise formal mathematical definitions. (Kirchhoff, 2010)

Any social process or system that can be conceptualized as a set of units and a set of lines connecting pairs of units can be studied as a social network. It can be constructed for any kind of social community and are studied because of the increasing interest in patterns of human interactions and the implications for the spread of information. (Newman, 2001, p. 1) Examples of social structures that have been studied as networks are friendship among children in a school, family relations among members of a social elite, shared board members of corporations, trade relations between countries, and hyperlinks between websites (De Nooy & Crothers, 2010)   
The tools of social network analysis help to describe the roles of individuals and small groups within the network. SNA measures are divided into individual and group-level measures (Sauer & Kauffeld, 2013) whereas the main individual / node-level measure is centrality. Centrality measures identify the most important vertices within a graph (e.g. the most influential person(s)). Several centrality measures are distinguished: Closeness centrality (sum of its distances to all other nodes) and betweenness centrality (number of times a node lays along the shortest path between two other nodes) (Barabási & Pósfai, 2016). Subgroup measures show how a network can be partitioned into communities (Hawe, Shiell, & Riley, 2004). Communities, smaller but more densely connected groups within a network have important influence on individual behaviour (Jenson, 2007).  
Following we define indicators describing the structure of a network and the role of nodes. Many more are described in the literature, but we will restrict ourselves to these ones used in this work.

**Degree**

The degree of a node is the number of connections it has. In a social network based on mails, the degree of an actor is the number of mails sent and received. (Ouyang & Reilly)

**Undirected / directed**

In a social network a pair of actors (a dyad) either can be adjacent (connected) or not, if no relation is existent between them. A network is called undirected if existing relationships between actors are symmetric. This means that no distinction between sender and recipient can be mad or the distinction is not important. Friendship networks are undirected due to the fact that if one person is befriended with another the connection is reciprocal. If the direction is not inherently symmetric, in the sense that each relationship involves distinct ‘sender’ and ‘receiver’ roles, graphs are called directed graphs. This may be the case with email networks as it is not necessary that if someone writes an email to another person that the receiver also writes one back. (Butts, 2008; Otte & Rousseau, 2002)

**Loop**

An edge from a vertex to itself is called as a loop. Networks which have no loops and which are not multiplex (i.e. do not allow duplicate edges) are said to be simple.

**Path**

A path between two nodes exists if they are connected and consistently reachable by a sequence of node pairs. The number of connections between distinct node pairs defines the length of a path or distance. If nodes are not distinct the sequence is not called path but trail and if either nodes or connections are distinct it’s called a walk.

**(Geodesic) Distance**

The shortest path connection any pair of nodes is the geodesic distance. In a directed network the path between two vertices can be different caused by the path direction.

**Subgraph**

Most networks contain different sub-modules, so called communities. Networks modularity is high if there are dense connections between nodes within a community but spare connections between different modules. A connected component is a subgraph in which any two nodes are connected to each other. (Newman, Barabási, & Watts, The structure and dynamics of networks, 2006, S. 1-8)

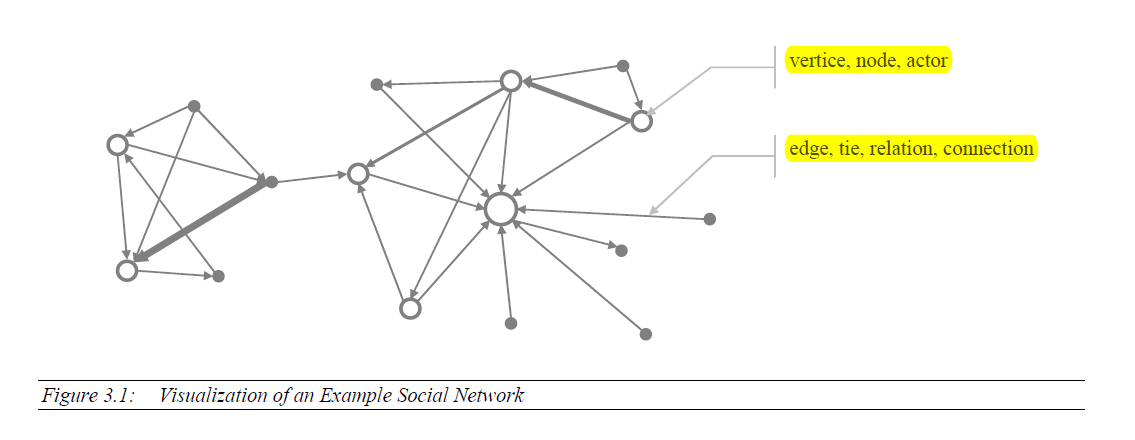
**Component**

A component is a special kind subset of a graph. The special characteristic of a component is, that there is a path between any node and any other one in this subset. If every node is directly connected to every other node in whole subgroup, the component it is said to be totally connected.

**Centrality**

Indicators measuring centrality identify the most important node in a network. Applications in social networks include identifying the most influential persons (Otte & Rousseau, 2002; Ouyang & Reilly, n.d.).

Figure2: Visualization of an Example Social Network

Source: Kirchhoff, 2010, p. 65

**Transitivity or clustering coefficient**

Transitivity measures the probability that the adjacent vertices of a vertex are connected. This is sometimes also called the clustering coefficient.

**Assortativity**

The assortativity coefficient measures the level of homophily of the graph, based on some vertex labelling or values assigned to vertices. If the coefficient is high, that means that connected vertices tend to have the same labels or similar assigned values

**Density**

Density of a graph is defined as the comparison between the number of existing connections divided by the number of possible connections between all actors within the social network. Density is an indicator for the general level of connectedness (Kirchhoff, 2010).

**Distance**

Distance is measured as the edge length of the shortest path between a pair of nodes. (Golubic, 2013, p. 2)

**Average path length**

The average path length in a graph is calculated as the sum shortest paths between all pairs of **vertices divided by the number of all paths.**

**Diameter**

Diameter describes the maximum of shortest distances between any two nodes in a network. If the graph is not connected the diameter is infinite. (Golubic, 2013, p. 2)

**Degree centrality**

Counts how many connections with acquaintances a node has. The more mails, the higher the degree centrality (Quelle: Node centrality in weighted networks: Generalizing degree and shortest paths)

**Betweenness centrality**

**The number of shortest paths from all vertices to all others a node is laying on is called betweenness centrality. Ascribe a high centrality to nodes connecting distinct communities together.** (Golubic, 2013, p. 2)

**Closeness centrality**

**Closeness centrality measures the mean of all shortest path to every existing node in the network. According to that its logic is different from the other two centrality measures, meaning nodes with low closeness centrality are more central and have a shorter way to reach all nodes.** (Golubic, 2013, p. 2)

**Degree distribution (power law)**

**The degree distribution of networks is an important topological property which impacts the dynamic of diffusion. The distribution of most social networks is skewed with many nodes having a low degree centrality and just a few having extraordinal many connections.**

**Communities**

**Detection of communities, one of the key properties of complex networks, is particular interesting. Community structures in networks describes groups of nodes that are more densely connected to each other than to nodes in other communities. These underlays the assumption that networks have natural divisions within it. Normally a node can only belong to one community, which is a simplification that not always holds in reality. Identifying communities in large networks is difficult and raised almost to a independent research field, filled with a lot of literature. (Quelle: Social Networks Analysis: Tools, Measures and Visualization)**

**Modularity**

**The modularity of a graph measures how good the division is, or how separated the different vertex types are from each other. It defined as:**

**m is the number of edges, Aij is the element of the A adjacency matrix in row i and column j, ki is the degree of i, kj is the degree of j, ci is the type (or component) of i, cj that of j, the sum goes over all i and j pairs of vertices, and delta(x,y) is 1 if x=y and 0 otherwise. (Quelle: igraph R manual pages: Modularity)**

**Louvain, Walktrap, Label propagation**

**In-block nestedness**

**FOCUS MORE ON COMMUNITY DETECTION**

Figure 3: [...]

[Abbildung einfügen]

Source: […]

# 

# 3 Research State & gaps

This chapter is two folded, part one is about email network analysis and the second part about connectivity patterns in networks.

## 3.1 Email network analysis

Many studies applying SNA measures to email networks study the Enron dataset, due to **the lack of large public email corpus (Zehnalova & Horak, 2015, p. 1). Diesner, Frantz, & Carley (2005) analysed the structural properties of a directed graph with weighted edges and identified key players in the crisis. Rowe, Creamer, Hershkop, & Stolfo (2007) reconstructed the social hierarchy based on a social score measure in an undirected graph while Gloor (2005) extracts evolution movies with temporal links and content analysis. Others studied the undirected network of Enron employees using network analytic measures (Chapanond, Krishnamoorthy, & Yener, 2005) or considered networks constructed by changing the minimum number of mails exchanged that an edge exists (Shetty & Rey, 2004)   
This thesis conducts a network analysis not on the Enron email corpus but on a new email dataset.**

## ****3.2 Connectivity patterns****

As mentioned in the part 2.1 Network Science, the discovery of recurring connectivity patterns stands at the very origin of network science. Ascertaining them is central to understand the underlaying microscopic (node-level) mechanisms of networks. Two macro-structural patterns emerge most prominent; nestedness and modularity. (Solé-Ribalta et al., 2018, p. 1)

**Community detection with modularity**

Community detection is a clustering problem with the goal to assign each node in the dataset to a community in a consistent and meaningful way (Shah & Zaman, n.d.). Whereas meaningful means, it helps to understand the daedal organization of complex networks (Fortunato, 2010; Lancichinetti, Fortunato, & Radicchi, 2008). Community detection is one of the most well-studied problems in the field of (social) network analysis. Modularity measures the strength of division of a network into communities as a scalar value between -1 and 1. High modularity implies dense internal connectivity within a community, but sparse connectivity between nodes belonging to different communities. (Barabási et al., 2006). Modular patterns and their identification are a characteristic predominantly in SNA but used in other fields as well (Coleman, 1964; Fortunato, 2010).   
Exact modularity optimization is computationally intractable so many approximation algorithms have been developed to reasonably identify communities in the structure of networks (Barabási et al., 2006; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Newman & Girvan, 2003). Without the intention of being exhaustive, the most widely used are: *Infomap*, which uses random walks; *Fastgreedy*, a greedy algorithm that optimises modularity score by pairing nodes for which it calculated the maximum improvement of modularity; Edge betweenness is getting communities by removing edges with high betweenness score; *walktrap*, a hierarchical clustering algorithm which assumes that short distance random walks stay within the same community; *labelpropagation* assigns nodes to the same community as the majority of its neighbours; *spinglass* which is based on the Potts model to find that edges should connect nodes of the same community; *leading* *eigenvector*, optimizes the modularity by using eigenvalues and eigenvectors of the modularity matrix and *Louvain* (multilevel), also a greedy algorithm, which assigns a different community to each node and then moving nodes to one of its neighbours with which it achieves the highest improvement in modularity. (Yang, Algesheimer, & Tessone, 2016, pp. 14f)   
Recent studies paying attention to the different community detection to a greater extent are the paper about community detection methods in social network analysis by Ahmad, Baharom, & Sapaat (2018) or the comparison between community detection algorithms on artificial networks by Yang, Algesheimer, & Tessone (2016)

**Nestedness**

Nestedness, first observed in biological networks, characterizes the geographical distribution of fauna in isolated, yet related landscapes (Patterson & Atmar, 1986). A complex system shows nestedness if the interactions of any node is a subset of the interactions from a node with a larger degree (Solé-Ribalta et al., 2018). Beyond biological systems nested properties occurs as well in several other, unipartite networks as inter-country trade relations (König, Tessone, & Zenou, 2014), inter-bank loans (Soramäki, Bech, Arnold, Glass, & Beyeler, 2007) and in the management field within inter-firm knowledge growth networks or disparate industry sectors (Tomasello, Napoletano, Garas, & Schweitzer, 2016).   
BINMATNEST (Binary matrix nestedness temperature calculator) from Rodriguez-Girones & Santamaria (2006), a further development on NTC (nestedness temperature calculator) from Atmar & Patterson (1993), the nestedness metric based on overlap and decreasing filling (NODF) (Almeida-Neto, Guimarães, Guimarães, Loyola, & Ulrich, 2008) and FCM (fitness-complexity metric) from (Tacchella, Cristelli, Caldarelli, Gabrielli, & Pietronero, 2012) are among the most widely used quantifying methods for nestedness. Recent studies in this field also potter at detecting new methods detecting nestedness in real-world networks, without the underlaying hypothesis that all nodes belong to a single nested component (Grimm & Tessone, 2017).   
The omnipresence of nestedness, discovery in fields despite ecology and its implications for stability in them, implicated research about the relation to other common network structures (Jonhson, Domínguez-García, & Muñoz, 2013; Lee, 2016; Solé-Ribalta et al., 2018).

**In-block nestedness**

A new debate among scholars started: Can modularity and nestedness co-exist? The two patterns appear incompatible, since the emerge from contractionary dynamics (cooperation and competition). A paper recently published by Solé-Ribalta et al. (2018) presents the concept of in-block nestedness. The concept and the corresponding method determine to what extent a network is composed of blocks (communities) whose internal connectivity exhibits nestedness. Different to other work they jointly consider both patterns and neither measures nestedness and modularity independently (Borge-Holthoefer, Baños, Gracia-Lázaro, & Moreno, 2017; Fortuna et al., 2010; Olesen, Bascompte, Dupont, & Jordano, 2007) nor operates sequentially (Flores, Valverde, & Weitz, 2013; Lewinsohn & Ina, 2006).

RESEARCH GAP

# 4 M-Industry & Project ONE

The following paragraph gives an overview about the organization the dataset is derived from.

## 4.1 Migros Group

Migros is the largest retailer and the largest private employer in Switzerland. The group comprises over 60 companies. Ten regional Cooperatives, each with an own management, are responsible for over one thousand stores. FMC (Federation of Migros Cooperatives), the central service provider of the Migros represents the group externally, is owned by the cooperatives. The operating business is divided into six departments. Beside the industry which together with wholesale forms the department four; HR, communications, culture and leisure time are summed up in department one. Marketing is department two, logistics and informatics department three, finance department five and retailing forms department six. All six departments are part of the general direction, which together with the assembly of delegates, the administration and statutory auditors outline the legal bodies of FMC.

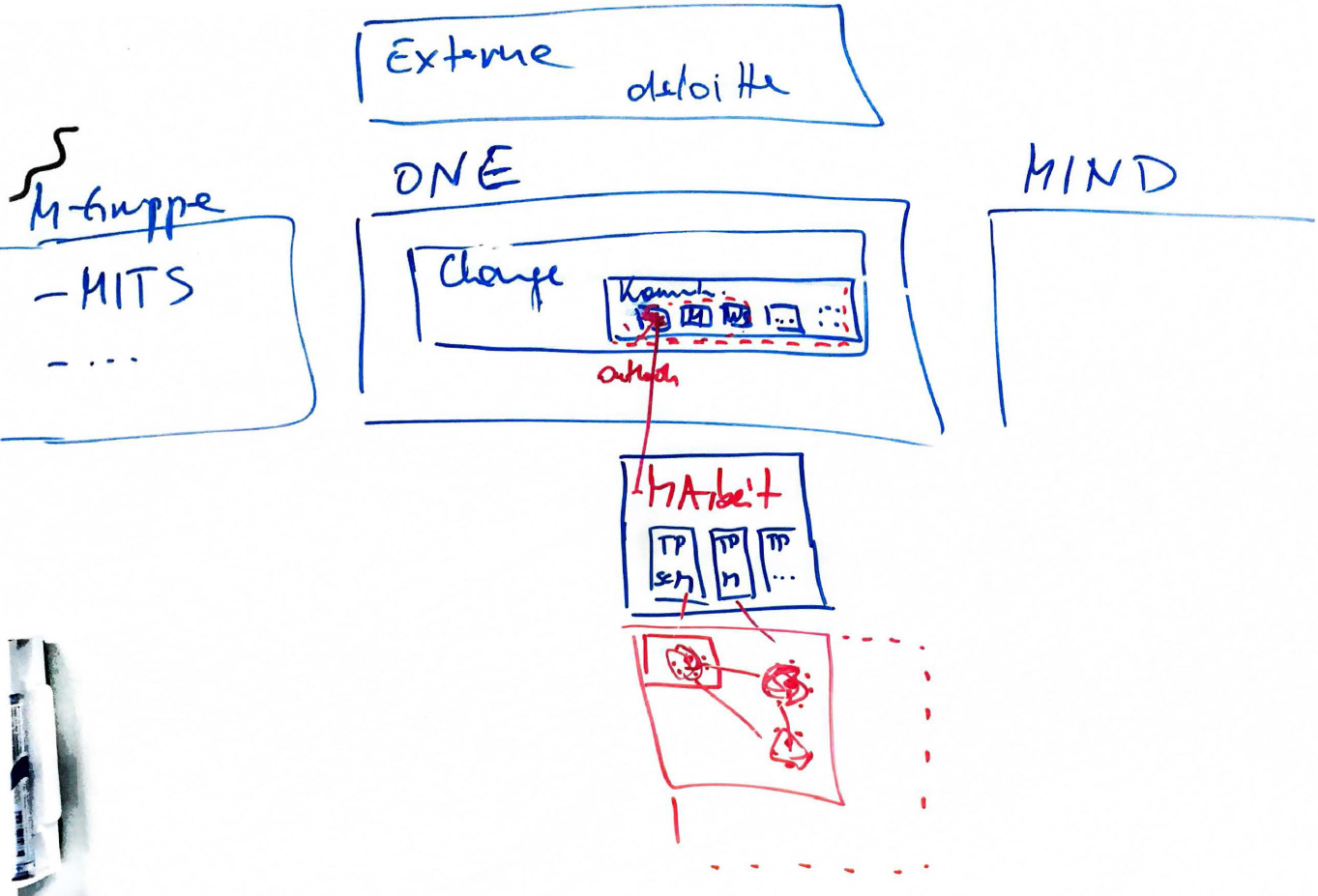
## 4.2 Migros Industry

As described in the previous section 4.1 Migros Group M-Industry is one out of six departments in the Migros Group. The M-Industry consists of 25 companies in Switzerland and eight companies abroad, which are mainly producing goods sold in the Migros supermarkets but as well in other stores (Denner, Volg, etc.) and overseas (United Kingdom, Germany, China, United States, France, etc.). Migros Industry sells over twenty thousand different food and non-food products, has more than fourteen thousand employees and total net sales of 6.5 billion. (Migros-Genossenschafts-Bund, 2017, 2018b, 2018a)

Within the M-Industry, more exactly among the 25 companies in Switzerland, which are all independent but affiliates of FME (Federation of Migros Cooperatives), efforts towards increase the competitive ability are rising following the goal to grasp the opportunities in foreign markets. The latest and biggest action to increase the efficiency and effectiveness is called ONE smart solution.

## 4.3 ONE

ONE is a change project launched in September 2016 with the goal to roll out a uniform ERP system (SAP Hana) and restructure the underlying processes towards centralization to exploit better the market power. Meaningless to mention that a project with a budged of over 50 million is strategically important for M-Industry and rather big in its organisation. The core team consists of 150 employees. Beside the core team almost 400 employees are part of so-called site groups and working part time in the project, beside remaining part of the original organisation in their “home” company. All team members, core team and site groups, are organized in thirteen subprojects. (Quelle: Intranet, Projekt ONE)   
Due to the scope and impact of ONE smart solution change management is a core task of the project organisation, beside the work on the processes and the modelling of them in the new ERP system. The PMO (Project Management Office) includes a dedicated person for communication, supporting the change activities and readiness within M-Industry. Within change management communication, inside the project team but as well as to the outside, plays a crucial role. This thesis is about the communication within and between the core team and the site groups.



Kommunikation: Outlook, Meetings, Workshops, Newsletter, …  
--> Konzentriert sich MA auf **Outlook** oder mehr von Kommunikation, sowie nur **projektinterne** Kommunikation oder auch nach aussen?   
--> Daten: nur Outlook, aber qualitative Abschätzung von Anteil  
--> Scope: nur projektinterne Kommunikation, wegen Datenschutz von Migros  
The case of ONE smart solution is particularly interesting because of its setting with at least three different layers of people (core team, thirteen subprojects and operational site groups).

# 5 Hypothesises

Based on the experience as an internal consultant in Migros Industry we have several hypothesises on what the result of the network analysis will be.

The communication within a subproject, a group of people dedicated to work on particular subfield of the project ONE, is expected to be dense. Typically, there is dense communication within a subproject, since members of a subproject are closely co-working on the same topic and originally come from similar business units in different firms and therefore speak a common language.

**Hypothesis 1:** A lot of communication within a subproject team

Co-workers originating from different business units often show differences in culture. Further, if people work in different subprojects, they deal with distinct topics. Consequently, we expect, between team members of different subprojects, weak interaction.

**Hypothesis 2:** Weak communications between different subproject teams

Assuming subproject membership defines the communities, these results would coincide with findings from community detection in network theory. As stated before, different clusters, in our case subprojects, are only joined by one or a few highly connected individuals but show dense connectivity between edges in the same community.

Weak communications between different subproject indicates a weak flow of information between two partial projects. This effect even is accentuated for site group members. Site group members are still engaged within the daily business of the company of origin. This results in spending less time for project one and showing little presence in the projects head office in Dietikon. We assume the time spent for the project ONE by site group members is focused on the particular subproject they’re assigned to, resulting in a lack of information about other subprojects.

**Hypothesis 3:** Lack of information for site groups about other subprojects

# 6 Concept & Method

The thesis uses existing methods on a new dataset, both described in the following paragraph.

## 6.1 Data

The email log is the main data. To estimate the proportion of communication represented in it, I also conducted a survey about communication in the project.

### 6.1.1 Email log

If you think about communication in business context, the first thing that comes to your mind is email and systematically collected data is the most reliable data. The IT department of Migros (MITS) stores emails in internal servers regardless of this thesis. This provides a way to extract relevant information without hurting personal rights and privacy. (Nitz, 2013, p. 50)

We use a dataset, which contains emails sent or received between ONE project members. The member list contains all names involved in the project organization. The information base comprises every single mail sent or received between core team and site group members during the time period from November 2016 until the end of June 2018. The dataset includes 595 people and over two million mails.   
Each individual has the attributes subproject, role and company. The whole project team comprises 18 different subprojects and ten different roles are possible – independent of each other. The 595 collaborators (originally) belong to 24 different companies.

The concerns regarding privacy are understandable and came up from ONE-managers, the IT and the legal department. In this work I address the question in two different ways. The dataset is fully anonymized by a third person so that only Migros possesses the key to verify who is the respective person behind the anonymous identity number. Nevertheless, the anonymization includes every critical aspect of an individuum such as subproject membership and role. This assures the results still can be aggregated by subproject or industry. Second, we don’t mine any content but only header information, which includes the sender, the receiver and time-stamp. The subject is only used to distinguish between private mails, which are excluded from the analysis, and business mails.

**Restrictions**

The collection contains only the communication within the project organization (core team and site group members). The organizational structure of Migros, with a variety of independent legal bodies, requires approval from all involved companies to use to communication towards the outside as well. The approval process is time consuming and would exceed the time limit of this master thesis.   
Moreover, steering committee members and the accompanying mails send and received, are deleted from the dataset.

### 6.1.2 Questionnaire

Interpersonal communication is not only constituted via email. To estimate the proportion captured by analysing the email connections, I conduct a survey among the affected persons. The questionnaire (Anhang XY) sent to all 595 persons tries to answer the question: “How much of the information flow happens via e-mail?”   
Even if the analysis of mail connections won’t represent the full truth, together with the questionnaire a good extrapolation becomes possible and we can see how much off the communication is included in the social network analysis.

## 6.2 Operationalization of the research question

To answer the question “How does network analysis help to map and understand the flow of communication in a project in real business environment?” node level measures – to understand the role of individuals – are used. Beside the average path length, the diameter and the degree distribution I analyse transitivity, the probability that an adjacent of a friend is my adjacent as well. Density gives back the portion of existing connections compared to the number of connections possible. Different centrality measures such as degree centrality, betweenness centrality, closeness centrality are used to identify the most influential individuals. Beside the above results I plot the whole network to get a visual impression of communication flow.

The hypothesises one to three are about community detection. Modularity measures exactly what the hypothesises are about. The measure is high if there are dense connections between nodes within a community (subproject) but spare connections between different modules. Based on the paper “A comparative analysis of community detection on artificial networks” from Yang et al. (2016) and because the in-block nestedness model from Solé-Ribalta et al. (2018) uses Louvain-method from Blondel et al. (2008) I use the Louvain- or also called multilevel-Method for community detection. Beside this modularity maximizing approach I also use *walktrap* and *labelpropagation* to check and ensure Louvain has the highest modularity value. By comparing the Louvain-communities based on the empirical communication data with the administrative people allocation into subprojects the hypothesises can be tested. If the modularity maximized communities are congruent with the subprojects hypothesis one and two are confirmed. If communities don’t match well with *subproject.type* the node attributes *sp1.type* and *company.type* may map better.   
For hypothesis one and two I only use the core team members. Each core team node is assigned to a subproject (sp1.type in the node table). Hypothesis three is about the site-group members which all have sp1.type zero, because they are not officially assigned to a subproject by the project management office (PMO). To answer hypothesis three, I first have to assign the SG-nodes to a subproject. I do this in the most intuitive way by checking for every node into which subproject it has the most existing edges and then assign it to the corresponding *sp1.type* itself. After that the same procedure as for hypothesis one and two is used to check weather site groups lack information about other subprojects. Hypothesis three predicts a high modularity value for site group members and congruent subprojects and communities.

The question about the understanding of connections between individuals is the question about predominant connectivity pattern. The two predominant connectivity patterns, nestedness and modularity, can occur interlinked (Solé-Ribalta et al., 2018). This finding challenges the understanding of the topology of social systems, calling for new models to explain the emergence of in-block nestedness. To test if interpersonal connections in the ONE dataset follow the IBN patterns I use the method from the paper of Solé-Ribalta et al. (2018).

To generate insights to improve information spreading I compare the SNA results with real events throughout the different project phases, trying to find meaningful commonalties. Comparing the communalities with the business experiences can lead to indications how improvements on communication can be done.

## 6.3 Execution

Centrality and most node level measures can be calculated for directed and weighted networks. The Louvain community detection algorithm demands networks to be undirected and the in-block nestedness algorithm from Solé-Ribalta et al. (2018) only runs on undirected and unweighted networks. To get comparable results along all analyses, all calculations are conducted on an undirected and unweighted network. Further I use undirected edges because the business’ focus of interest lies on who interacts with who but not in the distinction between sending and receiving mails. The correlation between in- and out-degree is 0.9 and reciprocity of communication in the project ONE dataset is 0.76.   
The calculations of the biggest component and assortativity mark exceptions and are calculated with the directed graph as well, due to different interpretation for undirected and directed versions of the network. Same for degree calculations of the nodes, which is conducted with the undirected weighted and the undirected unweighted graph due to the fact that the first represents mails and the second represents people connections.

The analysis is split into three parts and except the *DynSnap* partition and the IBN calculation I used the program *RStudio* along the *igraph* package for social network analysis.

### 6.3.1 Static network

To get an overview of the whole communication, time and date are ignored in this part. The network contains the whole dataset and shows who sends mails to who respectively who interacts with who.

First, we plot the whole undirected unweighted network as a graph and count the vertices and edges of which it consists. The basic *igraph* commands are used on the undirected and unweighted network to calculate basic measures of a networks: Transitivity, the average path length and the diameter, including the edges on it. The components function shows the number of strong and weak connected components together with the corresponding sizes. To catch the difference between strong and weak connected components the directed network is used, otherwise every existing connection would be strong. Assortativity for the weighted graph gives back if people sending or receiving a lot of mails mostly interact with people also communication a lot via email. The assortativity of the unweighted graph tests if high degree nodes interact mostly interact with other high degree nodes, whereas a high degree means interaction with many different people, not with many mails. Because of the different meaning of assortativity of simplified or not simplified networks, both are calculated.

Degree, betweenness and closeness centralization indicate whether the graph on global level is tightly organized around its most central point or not. Different centrality measures provide insights for the most central – sometimes called most important – individuals. The histogram for each of them shows the distribution. A comparison with normalized mutual information measure indicates if the most central node varies depending on the centrality used. I also compare the five most central vertices of each measure manually.   
The analysis on the interaction network, hence the unweighted network is used for centrality calculations. Nevertheless, it’s interesting to know how many mails are sent and who sends the most, therefor I recalculate degree centrality un-simplified network as well.

The *igraph* package includes commands for multi-level community detection, community detection based on propagating labels and via short random walks. Modularity helps to rate for which community detection algorithm the partition is best.   
The communities from the method with the highest modularity we then use to compare against the administrative clusters. We compare the empirical communities not only against the subproject membership but also against company membership and the role of the individuum in the project ONE. For hypothesis one and two only core team nodes (*r1.type* ≠ 1) are considered.   
Site group members have no subproject assigned to them. To test of hypothesis 3, they need to be assigned manually to a subproject. I do this by checking for every site group node into which subproject it has the most connections. Then the site group member gets assigned it to this subproject. The empirical communities based on communication are the compared to the assigned subproject, using the same method as for the core team members. To get a complete picture I also do it for the whole project team, core team and site-group members, even this has no meaning for the hypothesises. The code

### 6.3.2 Dynamic network

The collaboration network is dynamic and we have the time stamp of every mail. To represent the timely evolution, the dynamic networks normally are represented as a sequence of snapshots. Darst, Granell, Arenas, Gómez, & Saramäki (2016) elaborated an approach, which does not use constant but dynamic intervals. They propose a method able to detect evolutionary changes in the configuration. The code looks at similarity peaks between the event sets and it is publicly available (Darst, Franell, et al., 2016). I use their code to split the dataset into different snapshots to examine the network evolution in time. no input parameters or a priori assumptions that must be made before the method can begin. The input is simple, consisting of numeric event-ID (*Von* / *An* columns) and time. The method finds an initial intrinsic scale to the data, where each interval represents roughly the same amount of change.

After defining the intervals, the analysis of the system turns into an investigation of the slices. All analyses described in *6.3.1 Static network* are conducted for each snapshot of the network*.* A time series for each measure shows the evolution over time and makes a comparison possible.

### 6.3.3 In-block nestedness

The last part of the analysis concerns the overall connectivity pattern in the collaboration network. I test for in-block nestedness with the model from Solé-Ribalta et al. (2018). The code for IBN detection runs on Ubuntu and uses the Louvain algorithm for community detection. As the network is dynamic, I use the *DynSnap* snapshots to analyse time variation of in-block nestedness over time. For each of the eleven networks of different points of time I run the calculation fifty times and create the average out of it and plot the standard deviation as well. So, we can see the trend on a time series plot.   
Even if the algorithm already subtracts the IBN that can be accounted to randomness, I do the same analysis for two random networks with the same properties as the ONE-network. The two random networks are first the Erdös-Renyi created with e*rdos.renyi.game* command in igraph using the same amount of vertices and edges as in the original graph. The edges are chosen uniformly randomly from the set of all possible edges. The second random graph is with a given degree sequence. The configuration model connects the out-stubs of the edges together. The IBN for the two reference models should be zero or near to zero since, as mentioned above, the algorithm accounts for the random part of IBN.   
To verify if the time series for the ONE-network is significantly different to the curves for the Erdös-Renyi-Networks and the configuration model I perform a paired t-test on a 99% confidence level.

# 7 Results

Jede einzelne Kennzahl / Analyse was es ist und wieso gemacht und was das Resultat ist

## 7.1 Description of population & sample

### 7.1.1 Questionnaire

As completion rate of the survey is 30.6%, we received 216 complete answers from 705 visits (multiple visits possible). 478 visitors did only visit the site but not start the questionnaire and nine finished uncomplete. Out of the 227 visitor who begun the questionnaire only 9 didn’t finish.

Email rated as the most used communication tool (7.1/10), followed by personal interaction (5.7/10), phone (5.4/10), Skype calls and messages (4.9/10) and WhatsApp (3.4/10). Slack, Jira and Confluence are less used, even they are promoted by the projects managing office.   
The questionnaire further distinguishes between communication within the same subproject and communication between subprojects, as we also do in network analysis. In intra subproject communication 38.6% percent use mails at least on a daily basis while only 30.7% have face-to-face conversations every day. Phone and Skype are used daily by 18.2% respectively 15.4% of the population. The other means of communication only play a submerged role in the daily communication between members of the same subproject and altogether have a share less than 4%. By excluding face-to-face interactions, because they are hardly measurable, the results become even more clear: 60.4% of the indirect communication is email traffic, 21% phone calls, 11.1% via Skype and 7.6% via other tools.   
The communication to the outside of the own subproject shows a similar picture. Communication via mail still is the most popular. 33.4% use mails daily, 21.8% use face-to-face interactions and 17.7% phone calls. Skype is used every day by 13% and the other communication tools combined by 0.9%. Excluding direct interactions 65% of communication is email, 18% phone calls, 7.8% Skype and 9.2% via other communication channels.

Email is still the most important communication tool. It is used more often than every other tool and as well more often than face-to-face communication. A social network analysis based on mails captures a major part of the interaction in collaboration and the communication paths.

### 7.1.2 Social networks

The whole dataset includes 2’003’867 mails in a time span of 595 days or 85 weeks, amounts to an average of 3’367 mails per day (including weekends) or 23’575 mails per week. The temporal distribution in Figure 3 shows the number of mails per day. The recurrent local minimums are the weekends and the drops at the end respectively beginning of each year show the holiday effects during Christmas time and new year. We may also observe the effect of summer holidays in the year 2017 during the end of July and beginning of August. The peak in March 2018 is due to the start of the pilot process in segment 3 among the companies Aproz, Bina and Jowa where the first fit gap analysis took place.

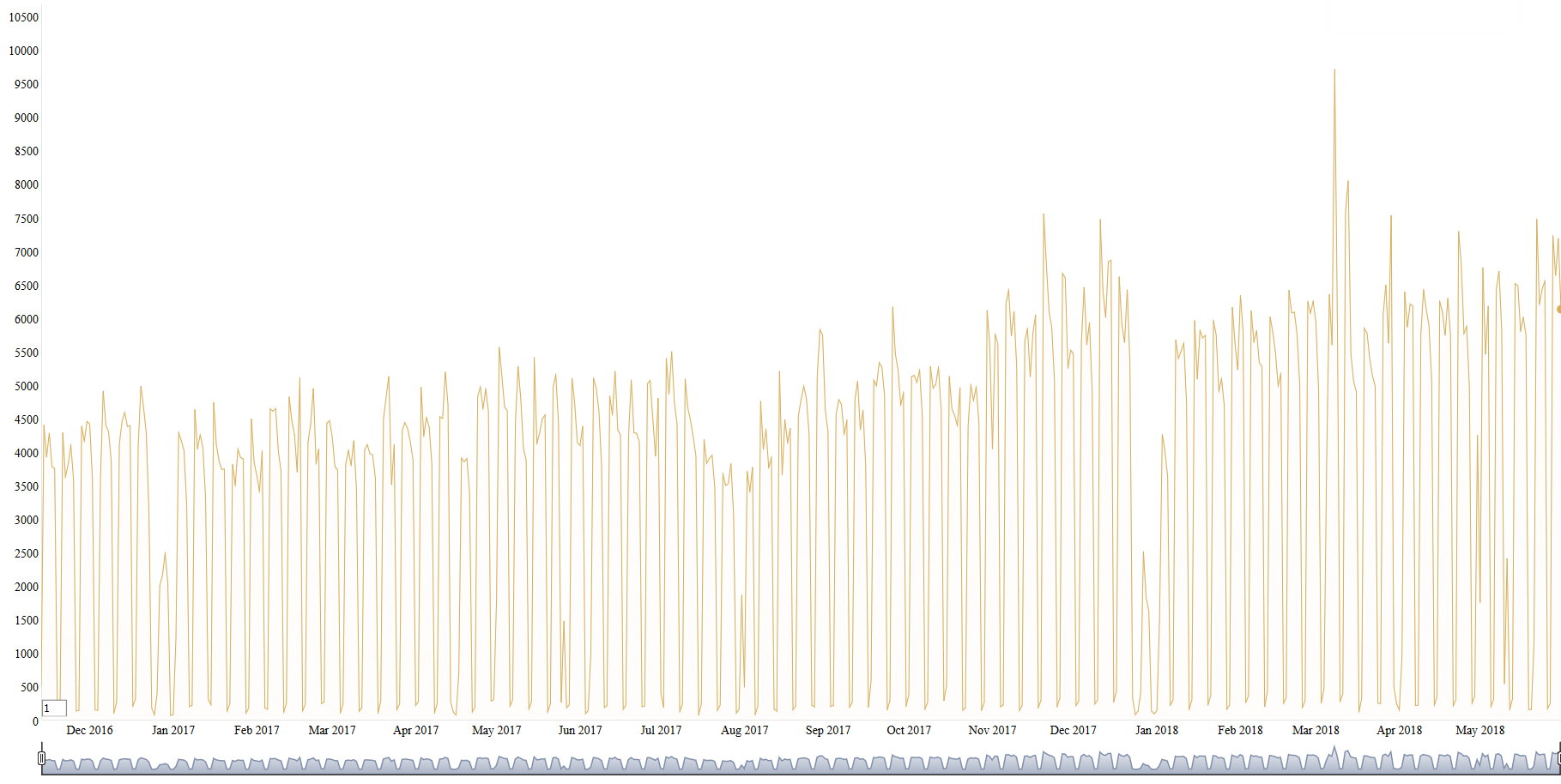


Figure 4 Mails per day during the period under consideration. The many sharp drops are the weekends on which less mails are sent.

The undirected and unweighted network contains 19’037 unique connections between the 595 individuals. Resulting in interactions to 64 different persons on average (average degree), because each interaction involves two nodes.

The *DynSnap* model divides the dataset into eleven slices (Figure 4). The timescales are: November 13th 2016 to January 31st 2017, January 31st 2017 to April 19th 2017, April 19th to June 13th 2017, June 13th 2017 to August 3rd 2017, August 3rd 2017 to September 14th 2017, September 14th 2017 to October 31st 2017, October 31st 2017 to December 6th 2017, December 6th 2017 to January 23rd 2018, January 23rd 2018 to April 1st 2018, April 1st 2018 to May 30th 2018 and May 30th 2018 to June 30th 2018. Figure 5 shows the dynamic network evolution. Each slice comprises between 5’365 and 9’613 interpersonal interactions resulting in 71’585 to 224’658 emails. Time intervals of several weeks appear reasonable for changes in email communication patterns (Darst, Granell, et al., 2016, p. 4). The intervals follow changes in event composition and not only the changes in event rate.

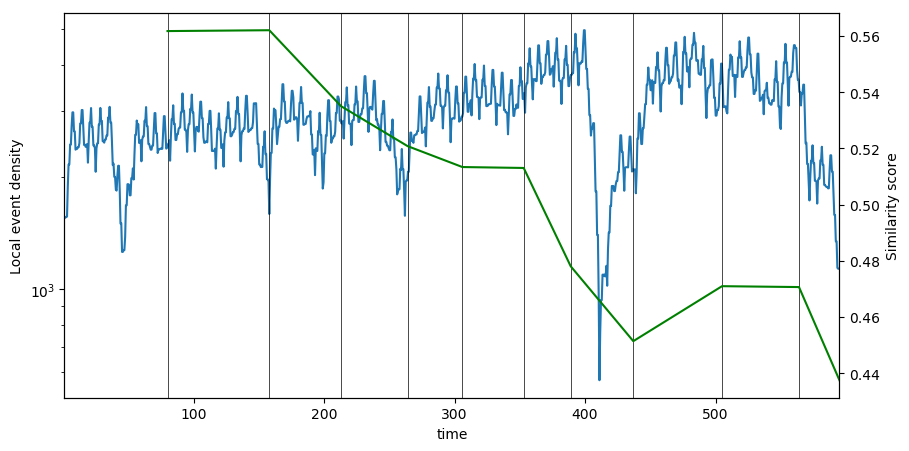


Figure 5 Slicing the ONE dataset. Each event is an email communication within the ONE project team. We can see that internal length reacts to the rate of change of events. The similarity reflects the rate of change of events at other times

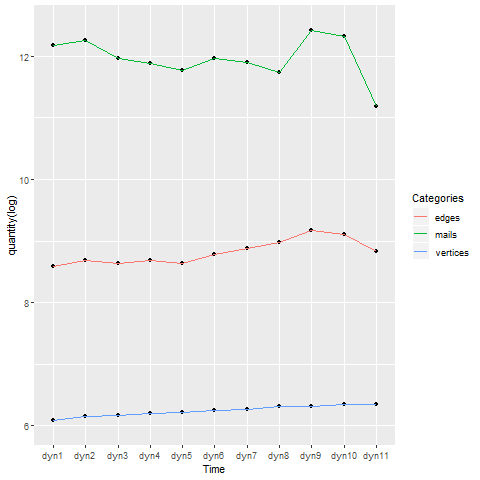


Figure 6 Dynamic network with DynSnap timescales. Number of vertices are raising constantly. Edges represent a connection between two individuals (number of vertices in undirected and unweighted network). Number of mails equals the number of edges in directed and weighted network).

## 7.2 Research questions & answers

Figure 6 is the graph representation of the analyzed network, containing all ONE project members, connected by a line if they interacted at least once via email in the time period from November 2016 until June 2018.

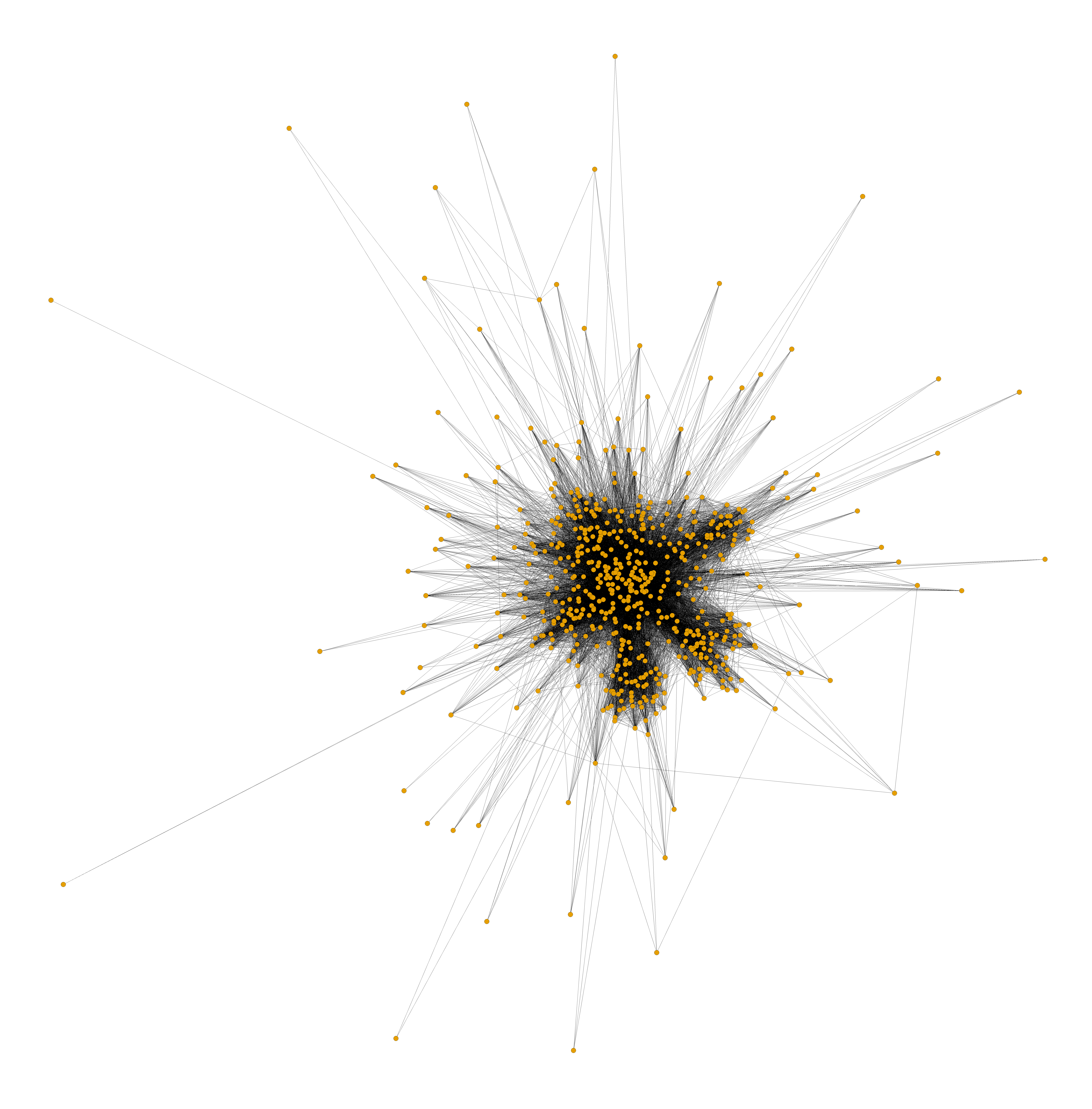


Figure 7 Complete ONE network (simplified). Including 595 vertices and 19'037 edges. Lines represent an interaction (at least one email) between two actors. Vertices placed on the plane using the force-directed layout algorithm by Reingold & Fruchterman (1991).

### 7.2.1 How does SNA help to map and understand the flow of communication?

Between the 595 actors there are 176’715 different (undirected) connections possible. The ONE network contains 19’037, resulting in an overall network density of 0.1%. The probability that adjacent vertices are connected (transitivity) is 0.38, meaning that the chances that a friend of my friend is a friend of mine as well are 38.6%. This is significantly higher than the global density, indicating a locally higher density. Assortativity, represents the tendency of similar vertices connect to each other. Since the value are positive, 0.2 for the simplified graph and 0.41 for the un-simplified, we can state that similar nodes (e.g. high degree nodes) more often connect to each other. The assortativity coefficient of the un-simplified networks (0.41) indicates, that people who communicate a lot via email do this with people also often communicating via email. The value for the simplified networks (0.2) shows the tendency of well-connected actors interacting with other well-connected people. The most basic and well-known node level measure is the average path length or mean distance. Within the project team the average path length is 1.9. On average every person can reach any other project team member (including site group members) in less than two steps. Short average path length in many real network leads to the concept of a small world, stating that every human being is connected by only six degrees of separation. Among the individuals in the ONE network the mean distance is even a lot shorter facilitating quick information transfer among the network and reducing costs. The longest path from an actor to another, measured by the diameter, is called size of the graph and equals to four for the ONE network.   
Figure 7 shows the evolution of the six measures. Assortativity raises, for both, the simplified and the un-simplified network. The longer the individuals are working together the more are well connected nodes interact among each other and the less they communicate with low degree nodes. Density, transitivity, the diameter and the average path length seem constant except for the timescales seven to ten (October 31st 2017 to May 30th 2018), where they trend towards a more densely connected network with higher transitivity and lower path lengths.

Freemans measure of graph centralization is an expression of how tightly the graph is organized around its most central point. Figure 8 shows the variation of the three centralization measures over time. Degree, betweenness and closeness centralization curve all show similar patterns. Constant during dyn1 to dyn6, raising from dyn7 until dyn9 and dropping to the original level for dyn10 and dyn11. For the time span under consideration closeness centralization is the highest in every *DynSnap* network, followed by degree centralization Betweenness centralization always is on a lower level than the other two. Considering the network of the whole timespan closeness centralization is 0.86, degree centralization 0.84 and betweenness centralization 0.24. The larger the centralization value is, the more likely it is that a single actor is quite central, with the remaining actors are considerably less central.

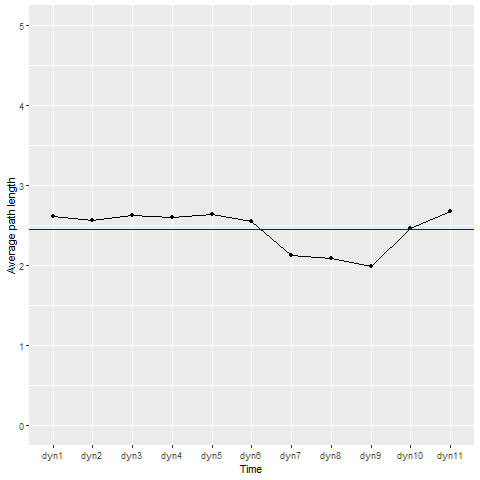
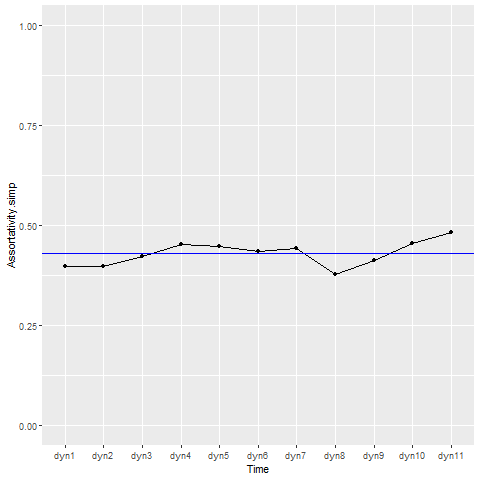
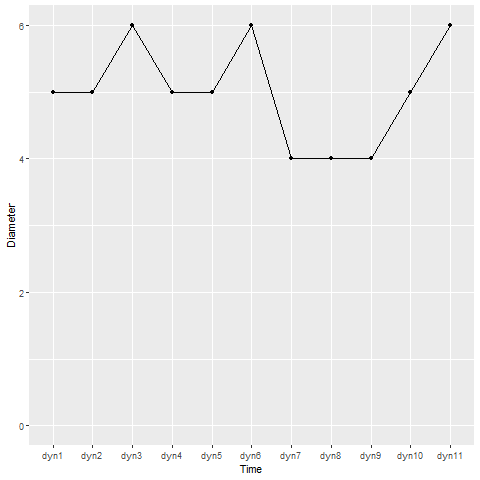
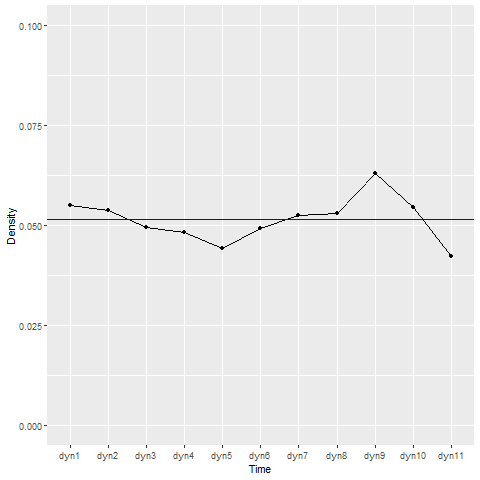
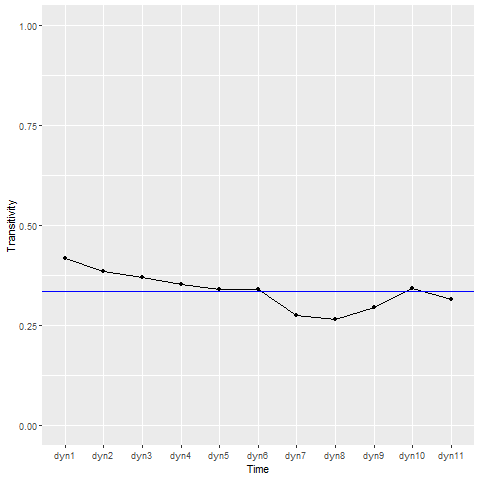


Figure 8 Node level SNA measures on the DynSnap timescales. Top row shows the evolution of assortativity in simplified (left) and un-simplified (right) network. Middle shows density (left) and transitivity (right). Bottom row is about path length, dimeter (left) and the average path length (right).

High degree centralization and closeness centralization score may indicate a node directly connected to most other nodes. This would imply similarly a high degree and a low average length of the shortest path between the node and all other nodes in the graph (closeness centrality). Figure 9 shows the distribution of degree centrality and closeness centrality. With both centrality measures one node has a far higher centrality value than the others. The betweenness centralization of 0.24 suggests that not a single node acts as gate keeper between most other nodes. (EV. BETWEENNESS DISTRIBUTION)

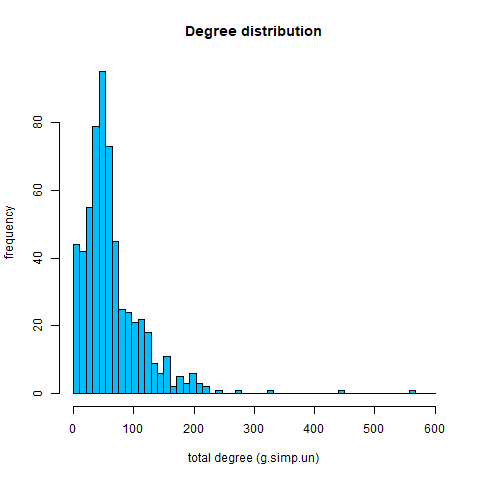
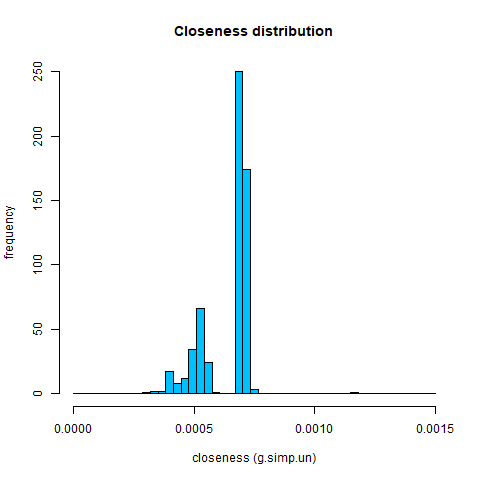
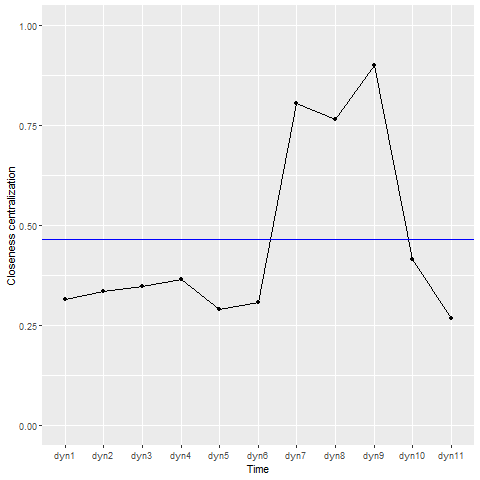
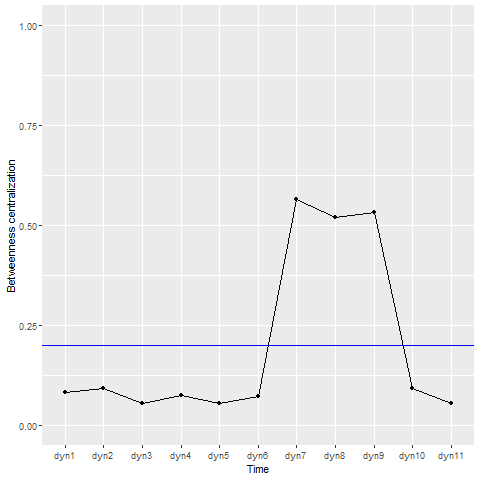
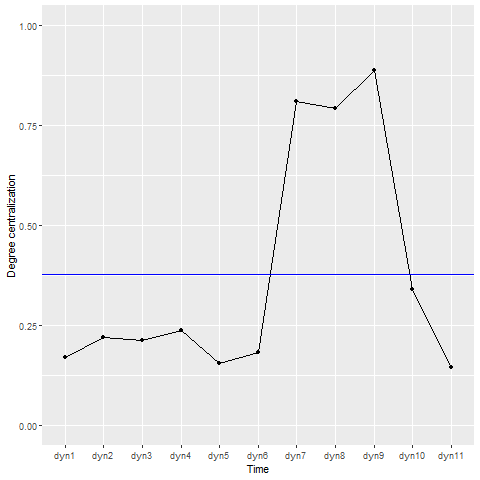


Figure 9 Centrality measures. Distribution of nodes. Degree centrality (left) and closeness centrality (right).

Figure 10 Centrality measures on DynSnap networks. Freeman's measures of centralization: degree (left), betweenness (middle) and closeness (right). expression of how tightly the graph is organized around its most central point. Three types of graph centralization to capture the various aspects of centralization.



### 7.2.2 H1 & H2: A lot of communication within subproject and weak communication between subprojects

The Louvain community detection algorithm has higher modularity than the *walktrap* and *labelpropagation* algorithm. The modularity value for the whole dataset is 0.70, indicating a good partition with dense communication within and sparse communication between the nine detected communities. The comparison between the empirically detected communities of the Louvain method and the administrative subproject belonging of core team members results in a mutual information measure of 0.23. NMI between Louvain and the company of origin is 0.48 and 0.18 compared to the role an actor has assigned.   
Figure 10 shows the mutual information measure of the Louvain communities compared to subprojects and companies of origin. For every snapshot at any time of the network evolution NMI for *company.type* is higher than for *subproject.type*, except for the last two. Over time the line comparing company to Louvain communities falls as the line for mutual information measure between subproject and communities rises.

Figure 11 Mutual information measure of Louvain communities compared to company of origin (left) and manually assigned subproject (right) for site group members.

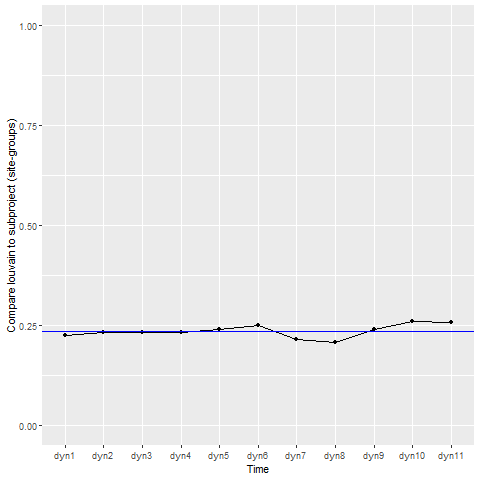
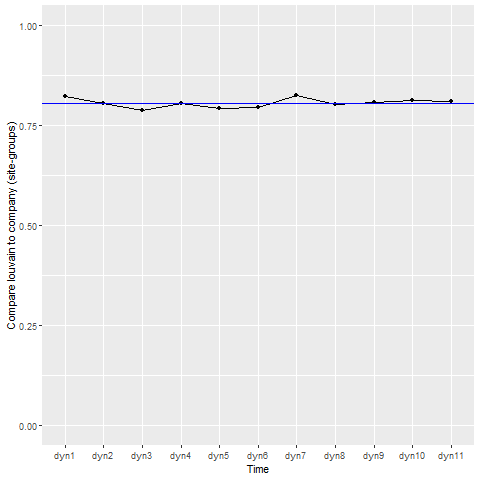
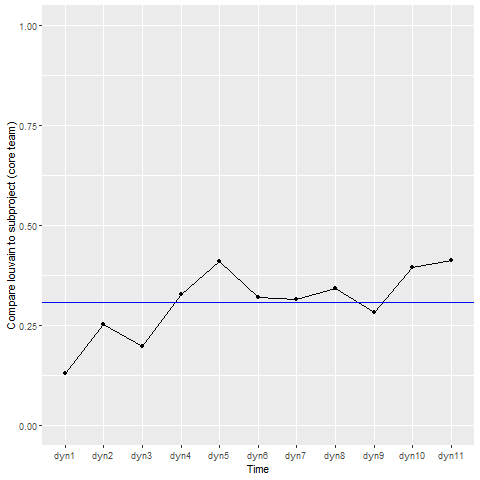
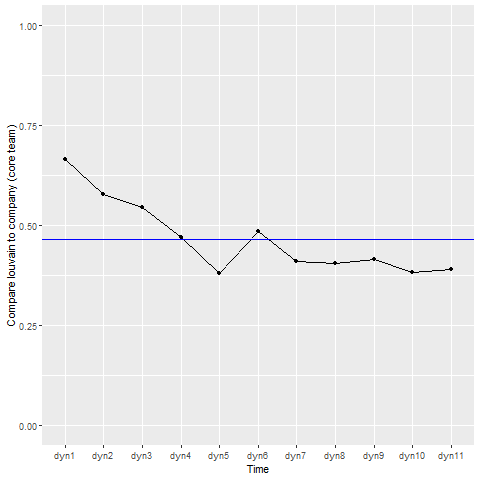


Figure 12 Mutual information measure of Louvain communities compared to company of origin (left) and assigned subproject (right) for core team members.



### 7.2.3 H3: Lack of information for SG about other subprojects

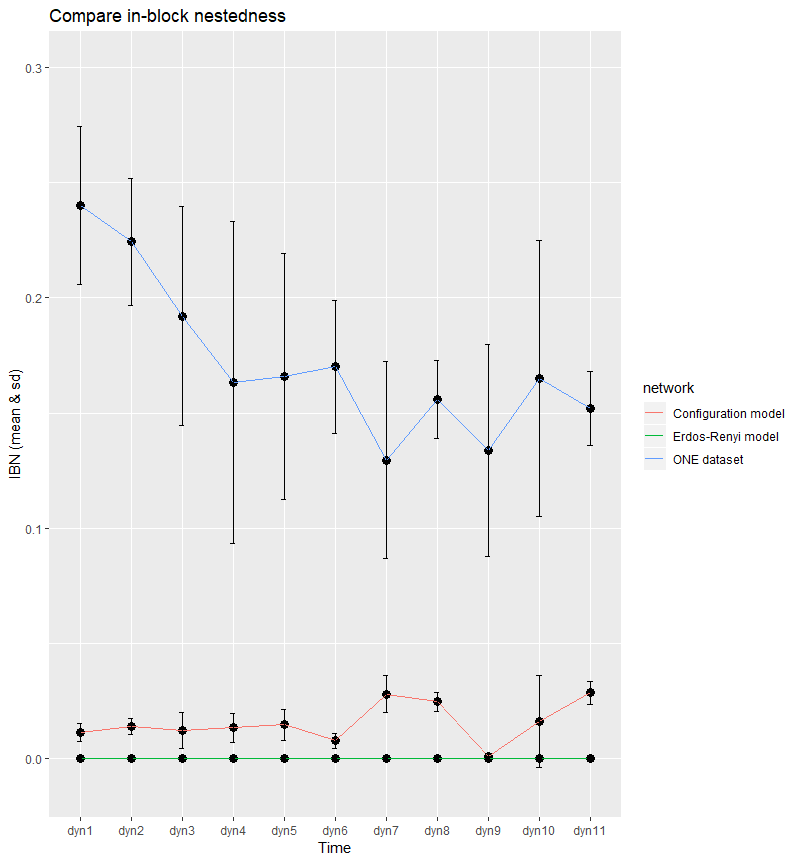
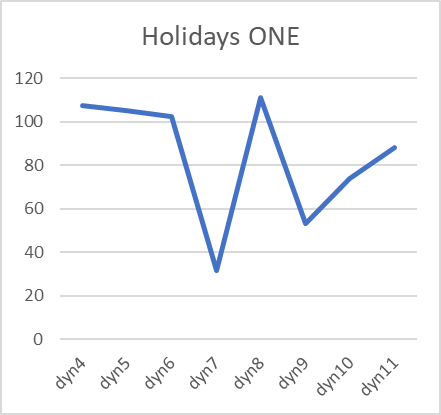
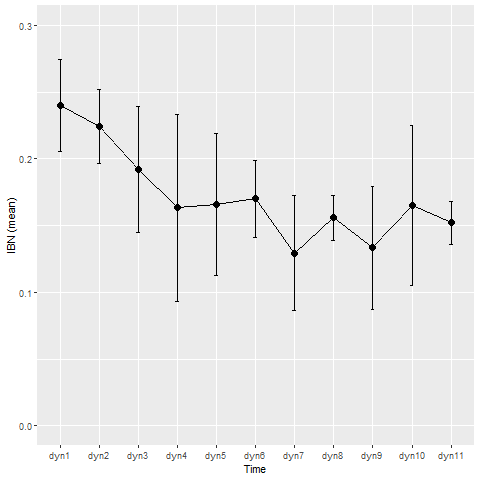
Using the same Louvain communities and comparing them for the site group, not the core team members, to the company of origin and the manually assigned subprojects, gives a mutual information measure of 0.22 for the subprojects and 0.86 for the companies. NMI of communities compared to roles is zero, because the role of the individuals is site-group member. As we can see in Figure 11 the curve for both comparisons are constant and no significant change of the NMI over time takes place.

Figure 13 In-block nestedness for the DynSnap timescales for the ONE network, Erdös-Renyi model and Configuration model. Values for the ONE dataset are significantly different from the two random model indication in-block nestedness.

### 7.2.4 How does SNA help to understand the connections between individuals?

Figure 12 shows the in-block nestedness derived with the model of Solé-Ribalta et al. (2018). IBN decreases from 0.28 to 0.16 between dyn1 and dyn4, stays around 0.16 – 0.18 until dyn6. Between dyn7 and dyn11 the IBN value alternates between every consecutive DynSnap network between 0.16 and 0.13. Figure 12 contains as well the IBN-lines for the two random networks generated with Erdös-Renyi and configuration model. These two lines are lower and close to zero for every DynSnap network. The line for all the ONE networks is significantly difference from zero at the 99% confidence level since the pairwise t-test gives back a p-value of 1.347e-08. Therefore, we can state, that in-block nestedness is present in the ONE collaboration network derived with emails as communication basis. The presence of IBN means the ONE team consists of communities with dense connection within and loose connection between and the connections within communities are nested.

### 7.2.5 Insights to help to improve communication



ALLES WAS AUF AUSGEDRUCKTEN BLAETTERN STEHT

## 7.3 Summary of results

SNA reveals that well connected individuals more often interact with other well-connected project team members (assortativity = 0.4, simplified network) and that communication via email often flows between two actors who are part of a lot of email traffic (assortativity = 0.6). Overall density of the network is relatively low, whereas on average individuals interact with around 10% of the possible communication partners within ONE. The dynamic analysis of the eleven DynSnap networks reveals lower transitivity, lower average path length and lower diameter but high centrality during the timescales dyn7 to dyn9 (October 31st 2017 to April 1st 2018). These measures may be indicators for increased top-down communication due to a higher need for coordination during the end of the Design project phase and the beginning of the Pilot & Common Core phase as well as the preparation for the first rollout wave.

The results described in the previous section reveal the presence of communities among the ONE core team. The communities detected by the Louvain algorithm contain a higher mutual information compared to the company the core team members originally come from than compared to the ONE subproject in which they belong. The dynamic analysis with the DynSnap networks show that NMI for communities and the company of origin decrease over time but increase comparing the empirical communities to subprojects. For the networks of the last two DynSnap intervals the mutual information is higher for the subproject core team members belonging to than for the company. Nevertheless, the NMI values are relatively low, indicating that Hypothesis 1 and Hypothesis 2 are not supported and that the communication within the subproject is not significantly denser than the communication between members of different subprojects.   
The community analysis for site group members reveals that information as well doesn’t mainly flow within subproject boundaries. The high NMI value comparing the Louvain communities and the company belonging suggests that there is a lack of communication between site group members origin from different companies within M-Industry.

Figure 12 proves the presence of in-block nestedness. Communication between ONE project members seem to support the recent finding that modularity and nestedness can occur jointly. ONE consists of internally densely connected blocks whose internal connectivity exhibits nestedness.

# 8 Conclusion

The emergence of interaction patterns between co-workers is a consequence of the dynamics in the underlaying communication network. The ultimate goal is to understand how information flows within the informal organization and whether this aligns with the perception in the organization.

We showed that the analysis of email communication captures a good amount of the overall communication among collaboration in business. Email stays the most important communication channel and the data stored in company owned servers allow social network analysis to quantify collaboration patterns and reveal information hidden in the informal structure of the interaction network.

The analysis of the ONE dataset has shown that communities are present. Low score for mutual information measure between subproject and company of origin compared to the membership to empirical community suggest that neither of both affects communication significantly for core team members. Therefore, we must state that the underlaying ground truth of the empirical communities is not found. One can think about other attributes than subproject, company shaping the community boundaries. Further investigations comparing other node attributes are necessary. Regardless of the absolute score, we saw that NMI comparing the company and community affiliation decreases over time, while NMI between subproject belonging and community increases. This suggests that initial communication boundaries, associated with the membership to the same company, break up with the time spend in the project organization. On the other hand, an individual communicates more within its subproject the longer it is not part of the home company anymore. These findings would make sense, since in the beginning of the project one would interact more closely with well known peers you worked together in the daily business of your home company. The more time you spend within the project organization and the more you collaborate within your subproject, the more communication flows within this community.   
Communities among site group members aren’t based on subproject affiliation neither. High mutual information between Louvain communities and the site group members company suggest that corporate affiliation mainly shapes interaction boundaries. The results show dense communication within members from the same community and weak communication between members of different communities. Considering that site group members only work part time for the project ONE and remain in their original company the other time, the results make sense. During the time spent in daily business communication mainly happens among co-workers from the same company.

The analysis of the general connectivity pattern shows the presence of in-block nestedness. This finding aligns with the determination of this structure in more social networks (Solé-Ribalta et al., 2018). The prove of this recent method on a new dataset suggest that previous work analysing email datasets may have overlooked this important feature when discussing collaboration networks.   
Nestedness as a structure on the mesoscopic network scale may indicate that each community has a team leader. The team leader aggregates all relevant information for its employees and distributes them among the needs and the hierarchy within the team. Noteworthy, it is not clear based on what attributes the communities are defined and if this attempt of explanation holds. Further work clarifying the ground truth behind the partition into blocks is suggested.

**IBN In Ferienzeit höher, da dann ev. die Personen die man kennt weg und man muss weg über subteam-leiter nehmen**

**Different project phases show different patterns in SNA (centrality and node level analysis)**

**Limitations**: This research was conducted using email communication data from a single organization. Hence any claims of generalizability are problematic. Field studies involving data from more organizations are needed before we can arrive at more definitive conclusions. Further research should compare actual face-to-face communications, telephone communications, letters, and memoranda along with electronic mail. Although employees included in the study constitute almost all the senior and middle level managements, it eventually excluded thousands of other employees. Again, this might be problematic as regards any claims of generalizability.

# 8 Recommendations for ONE

…

### Insights to help to improve communication

Comparison to real events

Increase in communication & connections (October 31st 2017 to May 30th 2018) comes together with the time of a different project phase (design) in which more coordination between the core team and the companies (site group members) is needed.

Centrality increases for dyn7 until dyn9. Eventually because of coordination from central position for the design phase

Communities based on company.type

EV. Discuss in last section (Recommendations for ONE

# Bibliography

[Achtung: Auf Formatierung achten 🡪 Hängende Abschnittsformatierung]

Bajaj Manu Bansal Guide, R., & Vullikanti, A. (kein Datum). Detection of communities in social networks. 20. Virginia: Virginia Tech.

Barabási, A.-L., & Pósfai, M. (2016). *Whole book\_Network Science.* Cambridge: Cambridge University Press.

Bendel, S. (2006). Das Kommunikationsverhalten von Unternehmensangehörigen. Hochschule für Wirtschaft Luzern.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*(4-5), 175-308.

Brandstädter, M., Grootz, S., & Ullrich, T. (2016). Kanäle der internen Kommunikation. In M. Brandstädter, S. Grootz, & T. Ullrich, *Interne Kommunikation im Krankenhaus, Erfolgskonzepte Praxis- & Krankenhaus-Management* (7 Ausg., S. 25-35). Berlin Heidelberg: Springer-Verlag.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1-9.

Dörfel, L., Mann, P., Boesler, B., Rosenzweig, U., & Herbst, D. (2016). *Trendmonitor Interne Kommunikation 2016.* School for Communication and Management [SCM] MPM Corporate Communication Solutions, Berlin.

Ehrlich, K., & Carboni, I. (2005). *Inside Social Network Analysis.* Boston College, Boston.

Huck-Sandhu, S. (2009). Interne Kommunikation in der Wahrnehmung von Mitarbeitern Ergebnisse einer Befragung. Hohenheim: Universität Hohenheim.

Mishra, N., Schreiber, R., Stanton, I., & Tarjan, R. (kein Datum). *Clustering Social Networks.* University of Virgina.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*(45), 167-256.

Newman, M. (2010). *Networks: an introduction.* Oxford: Oxford University Press.

Newman, M., Barabási, A.-L., & Watts, D. (2006). *The structure and dynamics of networks.* New Jersey: Princeton University Press.

Nitz, O. (2013). Instrumente und Techniken der Internen Kommunikation: E-Mails in der Internen Kommunikation (1 Ausg., Bd. 2). (L. Dörfel, Hrsg.) Berlin: scm c/o prismus GmbH.

Ahmad, J., Baharom, S., & Sapaat, M. A. (2018). Test case prioritization technique for event sequence test cases based on redundancy factor. *Journal of Theoretical and Applied Information Technology*, *96*(18), 6041–6052. https://doi.org/10.1166/asl.2011.1261

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D., & Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, *117*(8), 1227–1239. https://doi.org/10.1111/j.0030-1299.2008.16644.x

Atmar, W., & Patterson, B. D. (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, *96*(3), 373–382. https://doi.org/10.1007/BF00317508

Barabási, A.-L. (2014). Network Science Communities. *K-Mode*, *26*, 60. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjgxuGgr9bdAhUNqYsKHcPQAcIQFjAAegQIABAC&url=http%3A%2F%2Fbarabasi.com%2Ff%2F618.pdf&usg=AOvVaw1zXAxN9PeLvXFluV8UJ7ln

Barabási, A.-L., Newman, M., Watts, D. J., Barabasi, A.-L., & Watts, D. J. (2006). *The Structure and Dynamics of Networks*. *Information ResearchAn International Electronic Journal* (Vol. 11). New Jersey: Princeton University Press. Retrieved from https://books.google.ch/books?hl=de&lr=&id=xLfQCwAAQBAJ&oi=fnd&pg=PA8990&dq=the+structure+and+dynamics+of+networks&ots=Bmit5P2K4g&sig=Pl4bGqpJI7fYOUtBdawmjC4JrXk#v=onepage&q=the structure and dynamics of networks&f=false

Barabási, A.-L., & Pósfai, M. (2016a). *Network Science*. Cambridge: Cambridge University Press. https://doi.org/10.1098/rsta.2012.0375

Barabási, A.-L., & Pósfai, M. (2016b). *Network Science*. https://doi.org/10.1098/rsta.2012.0375

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 1–12. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borgatti, S. P., & Molina, J. L. (n.d.). *What Can Social Network Analysis Do for Management Consulting?* https://doi.org/10.1161/01.STR.32.1.139

Borge-Holthoefer, J., Baños, R. A., Gracia-Lázaro, C., & Moreno, Y. (2017). Emergence of consensus as a modular-to-nested transition in communication dynamics. *Scientific Reports*, *7*, 41673. https://doi.org/10.1038/srep41673

Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology*, *11*(1), 13–41. https://doi.org/10.1111/j.1467-839X.2007.00241.x

Chapanond, A., Krishnamoorthy, M. S., & Yener, B. (2005). Graph Theoretic and Spectral Analysis of Enron Email Data. *Computational and Mathematical Organization Theory*, *11*(3), 265–281. https://doi.org/10.1007/s10588-005-5381-4

Cheong, C., & Corbitt, B. (2009). A social network analysis of the co-authorship network of the Australasian Conference of Information Systems from 1990 to 2006. *ECIS 2009 Proceedings*. Retrieved from https://aisel.aisnet.org/ecis2009/6

Coleman, J. S. (1964). *Introduction to mathematical sociology*. Free Press of Glencoe.

Contreras Arias, C. (2017). The assets of the ‘poor’: rural indigenous livelihoods and their exemplar strategy to overcome scarcity - ‘Pos, nomás con lo “poco” que tenemos, todo lo que hemos logrado’. *University of St. Gallen, Business Dissertations*, (4652), 1–337. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=125256370&site=bsi-live

Darst, R. K., Franell, C., Arenas, A., Gómez, S., Saramäki, J., & Fortunato, S. (2016). GitHub - rkdarst/dynsnap: Dynamic time-slicing method. Retrieved 20 December 2018, from https://github.com/rkdarst/dynsnap

Darst, R. K., Granell, C., Arenas, A., Gómez, S., & Saramäki, J. (2016). Detection of timescales in evolving complex systems. *Nature Publishing Group*, *6*(July), 1–8. https://doi.org/10.1038/srep39713

De Nooy, W., & Crothers, C. (2010). *Historical developments and theoretical approaches in sociology*. *Social Network Analysis*. Eolss Publishers. Retrieved from https://books.google.ch/books?id=JoNZCwAAQBAJ&pg=PA377&lpg=PA377&dq=Social+network+analysis+(SNA)+focuses+on+the+structure+of+ties+within+a+set+of+social+ac-tors,+e.g.,+persons,+groups,+organizations,+and+nations,+or+the+products+of+human+activity+or+cogn

Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication Networks from the Enron Email Corpus “It’s Always About the People. Enron is no Different”. *Computational and Mathematical Organization Theory*, *11*(3), 201–228. https://doi.org/10.1007/s10588-005-5377-0

Flores, C. O., Valverde, S., & Weitz, J. S. (2013). Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal*, *7*(3), 520–532. https://doi.org/10.1038/ismej.2012.135

Fortuna, M. A., Stouffer, D. B., Olesen, J. M., Jordano, P., Mouillot, D., Krasnov, B. R., … Bascompte, J. (2010). Nestedness versus modularity in ecological networks: Two sides of the same coin? *Journal of Animal Ecology*, *79*(4), 811–817. https://doi.org/10.1111/j.1365-2656.2010.01688.x

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174. https://doi.org/10.1016/J.PHYSREP.2009.11.002

Freeman, L. C. (2004). *The development of social network analysis*. *Document Design* (Vol. 27). https://doi.org/10.1016/j.socnet.2005.01.010

Garton, L., Haythornthwaite, C., & Wellman, B. (2006). Studying Online Social Networks. *Journal of Computer-Mediated Communication*, *3*(1), 0–0. https://doi.org/10.1111/j.1083-6101.1997.tb00062.x

Gloor, P. A. (2005). Capturing Team Dynamics through Temporal Social Surfaces. In *Ninth International Conference on Information Visualisation (IV’05)* (pp. 939–944). IEEE. https://doi.org/10.1109/IV.2005.30

Golubic, K. (2013). Discovering Value in Academic Social Networks: A Case Study in ResearchGate. In *Proceedings of the ITI 2013 35th International Conference on INFORMATION TECHNOLOGY INTERFACES* (pp. 3–8). Zagreb: University Computing Centre - SRCE. https://doi.org/10.2498/iti.2013.0566

Grimm, A., & Tessone, C. J. (2017). Analysing the sensitivity of nestedness detection methods. *Applied Network Science*, *2*(1), 37. https://doi.org/10.1007/s41109-017-0057-9

Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: how &quot;out of control&quot; can a randomised controlled trial be? *BMJ (Clinical Research Ed.)*, *328*(7455), 1561–1563. https://doi.org/10.1136/bmj.328.7455.1561

Himelboim, I. (2017). Social Network Analysis (Social Media). In *The International Encyclopedia of Communication Research Methods* (pp. 1–15). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118901731.iecrm0236

Jenson, J. M. (2007). Neighborhood and Community Effects on Individual Behavior and Social Functioning. *Social Work Research*. Oxford University Press. https://doi.org/10.2307/42659901

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654. https://doi.org/10.1038/35036627

Jonhson, S., Domínguez-García, V., & Muñoz, M. A. (2013). Factors Determining Nestedness in Complex Networks. *PLoS ONE*, *8*(9), e74025. https://doi.org/10.1371/journal.pone.0074025

Kirchhoff, L. (2010). *Applying Social Network Analysis to Information Retrieval on the World Wide Web: A Case Study of Academic Publication Space.* *University of St. Gallen, Business Dissertations*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=82753382&site=ehost-live

König, M. D., Tessone, C. J., & Zenou, Y. (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics*, *9*(3), 695–752. https://doi.org/10.3982/TE1348

Kunc, M. (n.d.). *Strategic analytics : integrating management science and strategy*.

Lancichinetti, A., Fortunato, S., & Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. https://doi.org/10.1103/PhysRevE.78.046110

Lee, S. H. (2016). Network nestedness as generalized core-periphery structures. *Physical Review E*, *93*(2), 022306. https://doi.org/10.1103/PhysRevE.93.022306

Lewinsohn, T. M., & Ina, P. (2006). Structure in plant animal interaction assemblages /, *1*.

Mariani, M. S. (2017). *Introduction to Network Theory*. Zurich: University of Zurich.

Migros-Genossenschafts-Bund. (2017). Zahlen und Fakten - M-Industrie. Retrieved 12 June 2018, from http://www.mindustry.com/de/ueber-uns/uebersicht/zahlen-und-fakten

Migros-Genossenschafts-Bund. (2018a). Organisation & Struktur | Migros Geschäftsbericht 2017. Retrieved 12 June 2018, from https://report.migros.ch/2017/governance/organisation-struktur/

Migros-Genossenschafts-Bund. (2018b). Über uns - M-Industrie. Retrieved 12 June 2018, from http://www.mindustry.com/de/ueber-uns/uebersicht

National Research Council (U.S.). (2005). *Network science*. National Academies Press.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. https://doi.org/10.1073/pnas.021544898

Newman, M. E. J., & Girvan, M. (2003). Finding and evaluating community structure in networks, 1–16. https://doi.org/10.1103/PhysRevE.69.026113

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, *104*(50), 19891–19896. https://doi.org/10.1073/pnas.0706375104

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441–453. https://doi.org/10.1177/016555150202800601

Ouyang, F., & Reilly, C. (n.d.). Terminology - Social Network Analysis. Retrieved 30 November 2018, from https://sites.google.com/a/umn.edu/social-network-analysis/terminology

Patterson, B. D., & Atmar, W. (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, *28*(1–2), 65–82. https://doi.org/10.1111/j.1095-8312.1986.tb01749.x

Reingold, T., & Fruchterman, E. (1991). Graph Drawing by Force-Directed Placement. *Software, Practice and Experience*, *21*(11), 1129–1164. Retrieved from http://www.stat.cmu.edu/~brian/780/bibliography/00 layout and rendering/fruchterman-reingold.pdf

Rodriguez-Girones, M. A., & Santamaria, L. (2006). A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography*, *33*(5), 924–935. https://doi.org/10.1111/j.1365-2699.2006.01444.x

Rowe, R., Creamer, G., Hershkop, S., & Stolfo, S. J. (2007). Automated social hierarchy detection through email network analysis. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis - WebKDD/SNA-KDD ’07*, 109–117. https://doi.org/10.1145/1348549.1348562

Sauer, N. C., & Kauffeld, S. (2013). Meetings as Networks: Applying Social Network Analysis to Team Interaction. *Communication Methods and Measures*, *7*(1), 26–47. https://doi.org/10.1080/19312458.2012.760729

Shah, D., & Zaman, T. (n.d.). Community Detection in Networks : The Leader-Follower Algorithm, 1–8.

Shetty, J., & Rey, M. (2004). The Enron email dataset database schema and brief statistical report. *Information Sciences Institute Technical Report, University of Southern California*, *4*(1), 120–128.

Solé-Ribalta, A., Tessone, C. J., Mariani, M. S., & Borge-Holthoefer, J. (2018). Revealing in-block nestedness: Detection and benchmarking. *Physical Review E*, *97*(6). https://doi.org/10.1103/PhysRevE.97.062302

Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., & Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and Its Applications*, *379*(1), 317–333. https://doi.org/10.1016/J.PHYSA.2006.11.093

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., & Pietronero, L. (2012). A New Metrics for Countries’ Fitness and Products’ Complexity. *Scientific Reports*, *2*(1), 723. https://doi.org/10.1038/srep00723

Tomasello, M. V., Napoletano, M., Garas, A., & Schweitzer, F. (2016). The rise and fall of R&amp;D networks. *Industrial and Corporate Change*, *26*(4), dtw041. https://doi.org/10.1093/icc/dtw041

Tran, M., & Khaw, S. (2006). Enron | Business | The Guardian. Retrieved 8 November 2018, from https://www.theguardian.com/business/2006/jul/06/corporatefraud.enron

Welch, M., & Jackson, P. R. (2007). Rethinking internal communication: a stakeholder approach. *Corporate Communications: An International Journal*, *12*(2), 177–198. https://doi.org/10.1108/13563280710744847

Wellman, B., & Berkowitz, S. D. (1988). *Social structures : a network approach* (1st ed.). Cambridge: Cambridge University Press. Retrieved from https://books.google.ch/books?id=XIw4AAAAIAAJ&dq=B.+Wellman+and+S.D.+Berkowitz,+Structural+analysis+in+the+social+sciences+2:Social+structures:+a+network+approach&lr=&hl=de&source=gbs\_navlinks\_s

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, *6*(March). https://doi.org/10.1038/srep30750

Zehnalova, S., & Horak, Z. (2015). Email Conversation Network Analysis : Work Groups and Teams in Organizations, 1262–1268.

(2016). *Trendmonitor Interne Kommunikation 2016.* School for Communication and Management [SCM] MPM Corporate Communication Solutions.

Weitere verwendete Quellen:

[z.B. nicht-öffentliche Statistiken, Interviews, etc.]

* […]
* […]

Appendix

UMFRAGE ONE

SKRIPT R

Appendix A: […]

…

Appendix B: […]

…

Appendix C: […]

…

Eidesstattliche Erklärung

I hereby declare that this thesis has been composed by myself autonomously and that no means other than those declared were used. In every single case, I have indicated parts that were taken out of published or unpublished work, either verbatim or in a paraphrased manner, as such through a quotation. This thesis has not been handed in or published before in the same or similar form.

Zurich, 29/12/2018 Denis Krebs