WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

University of Münster
School of Business & Economics

Joschka Andreas Hüllmann

Measurement of Social Capital in Enterprise Social Networks: Identification and Visualisation of Group Metrics

Master Thesis

at the Chair for Interorganisational Systems
(Westphalian Wilhelms-University, Münster)

in collaboration with the Chair for Business Information Systems
(University of Sydney Business School, Sydney)

Principal Supervisor:	Prof. Dr. Stefan Klein
Associate Supervisor:	Prof. Dr. Kai Riemer
Presented by:	Joschka Andreas Hüllmann
	huellmann@uni-muenster.de
Submission:	1st March 2017

Contents

Figures

Tables

Abstract

Social Capital describes the value of social structures in a social network. Effects of Social Capital, which include the improvement of knowledge sharing, collaboration and cooperation, can be observed in Enterprise Social Networks. Thus organisations are increasingly looking to employ Enterprise Social Networks platforms to facilitate these effects. The generated data on such platforms can be analysed to measure Social Capital.

The measurement of Social Capital is achieved by operationalising Social Capital with the help of Social Network Analysis. It can be explained as the analysis of relationships between individuals and groups within a network. It is based on graph theory and represents individuals or groups as nodes and their interactions as edges. Typical goals of such an analysis include the identification of influencers, leaders and high-performers in an organisation.

This thesis proposes to represent social groups as collective actors to utilise already existing Social Network Analysis metrics. The metrics are collected in a metric repository which contains the calculation schemas and interpretations. The schemas formally describe how a metric can be calculated and the interpretations discuss the values of the metrics against the backdrop of the Social Capital theory.

The metric repository is used as basis for the development of a prototypical analysis platform. The platform calculates all the metrics for a social network based on a real-world dataset and visualises them in a website. The identified metrics, their visualisation and the applicability of the group model approach are discussed.

1 Introduction

Social Network Sites are in everyday use with facebook hitting the one billion mark of users[1]. They facilitate communication and enable users to interact with others by writing text messages (McAfee 2006). A Social Network Site (SNS), that is running within an organisation, is called an Enterprise Social Network (ESN). By employing Enterprise Social Networks, organisations look to create Social Capital (Fulk and Yuan 2013; Leonardi et al. 2013) and try to improve expertise and knowledge sharing (Ellison et al. 2015). Social Capital describes the value of social structures and interactions for actors to achieve their goals within an organisation (Riemer et al. 2015; Portes 1998). Supporting these social structures and interactions can result in improved knowledge sharing and collaboration, which leads to higher worker productivity (Mäntymäki and Riemer 2016). Thus organisations are increasingly interested in adopting such technologies (DiMicco and Millen 2007). By adopting Enterprise Social Network technologies, organisations generate Enterprise Social Network data that can be analysed.

To analyse such data and measure Social Capital from it, I operationalise Social Capital with the help of Social Network Analysis. Social Network Analysis can be explained as the analysis of relationships between individuals, organisations and groups within the network (Scott and Carrington 2014; Stieglitz et al. 2014). It is based on graph theory and represents individuals or groups as nodes and their interactions as edges (Hacker et al. 2015). While Social Network Analysis is originally applied to Social Networking Sites (Viol and Hess 2016), it is increasingly adopted by researchers to analyse Enterprise Social Networks. Typical goals of such an analysis include the identification of influencers, leaders and high-performers in the organisation (Berger et al. 2014; Riemer et al. 2015) and the location of knowledge in the organisation (Stieglitz et al. 2014). Such research has been conducted on the relationships of individuals, but research on the relationships of groups is still lacking (Ellison et al. 2015).

By defining a group as a collective actor of individuals (Riemer 2005, p. 111; Borgatti et al. 1998), I am able to transfer research on individuals to groups and utilise Social Network metrics that are developed for individuals (Viol and Hess 2016; Hacker et al. 2015; Riemer et al. 2015). This allows me to look from the external perspective on Social Capital (Riemer 2005, p. 89), which considers the relationships between individuals and a group, by using ego-centric Social Network metrics (Scott 2012, pp. 30f). The internal perspective of Social Capital (Riemer 2005, p. 92) considers the relationships within a group and is analysed by using global Social Network metrics (Scott 2012, pp. 33f). I select relevant and non-redundant metrics to compare the internal with the external per-

1 https://newsroom.fb.com/company-info/ (accessed 2016-10-15).

spective of Social Capital. Based on the complementary nature of both perspectives (Burt 2001, p. 48), the metrics are linked to a group's activity, performance, engagement and health.

The selected metrics are utilised to create a reusable metric repository which includes the interpretation of each metric and its calculation schema. It can be used as a resource for developing software systems that analyse the Social Capital of Enterprise Social Networks.

Based on the metric repository I build a prototype that calculates group metrics from Enterprise Social Network data. The data of the prototype is based on a real-world Enterprise Social Network data set provided by Swoop, who specialises in Enterprise Social Network analytics[2].

The analysis is implemented in various automated scripts and the results are exposed via a REST API. A frontend consumes these results and provides a visualisation of the metrics on a website. Such a visualisation makes the results of an analysis accessible to the average user, so they can "*understand [the result and] extract value from it*" – as mentioned by Google's chief economist Hal Varian[3] in 2009 (Kaisler et al. 2013, p. 996; Gantz and Reinsel 2011, p. 6). The website also serves as a knowledge hub informing users about the metric repository and the theoretical underpinnings.

This thesis first discusses the theoretical background with regards to Enterprise Social Networks, Social Capital and Social Network Analysis. It is followed by the explanation of my research approach and the description of Swoop's dataset. In section 4 the metric repository is proposed and in section 5 the technical and visual software design is reviewed. The discussion part is concerned with the strengths and limitations of the metric repository and the prototypical implementation. The thesis ends with the conclusion and the appendix which contains a full list of all metrics and SQL calculation schemas, the database structure and screenshots of the website.

[2] http://www.swoopanalytics.com/ (accessed 2016-10-15).
[3] http://www.mckinsey.com/industries/high-tech/our-insights/hal-varian-on-how-the-web-challenges-managers (accessed 2016-10-18).

2 Background

In Burt (1992, 2001) the author proposes a theory of maximum group performance in organisations. Based on Burt's theory, my goal is to measure group performance in Enterprise Social Networks. I operationalise the Social Capital theory with the help of Social Network Analysis methodology and apply established metrics to my Enterprise Social Network data set provided by Swoop. Before the theoretical background is discussed, the following terms are defined based on Riemer (2005, pp.81-85).

An Enterprise Social Network is an information system, which hosts a social network and thus acts as a social platform (Riemer 2005). A Social Network consists of its actors (users) and social relationships between actors. While it is formed through a social platform, the Social Network is not bound to a particular platform (Riemer 2005). Actors can be individuals or collectives (groups) (Borgatti et al. 1998; Riemer 2005). A relationship is formed through textual interactions over time (e.g. posts, comments) between actors in the network (Granovetter 1992; van Dijk 1997).

2.1 Enterprise Social Networks

In the following I shortly argue for the relevancy of Enterprise Social research and discuss prior research. The differences between Social Network Sites and Enterprise Social Networks are pointed out. The major components and the goals of using such technology are discussed, followed by the functionality exposed to the users. Inferred from the functionality, I describe the effects of Enterprise Social Networks on an organisation and its impact on Social Capital.

2.1.1 Relevancy

As of 2016 thousands of Social Network Sites of different kinds exist (Richter and Riemer 2009; Richter et al. 2011). They are being adopted and used throughout all demographics according to McClard and Anderson (2008) and Stieglitz et al. (2014). While Social Network Sites are popular among young demographics and used for their interactions with personal friends (Richter et al. 2011; Vie 2008), their usage in business and politics is increasing (Stieglitz et al. 2014). Social Network Sites get more attention from organisations (Richter et al. 2011; Richter and Riemer 2009). Instead of participating in public Social Network Sites, organisations are looking to deploy their own internal Social Network Sites (DiMicco and Millen 2007), which are called Enterprise Social Networks (McAfee 2006). The primary reasons for deploying Enterprise Social Networks within organisational boundaries include knowledge exchange and collaboration purposes (Richter et al.

2011). Since the organisation is in control of the network, it can analyse all generated data.

2.1.2 Prior Research

Enterprise Social Networks are still a small domain of research compared to Social Network Sites, but insights from SNS research can be appropriated (Richter et al. 2011). Early research on Social Network Sites was conducted in the domain of computer supported cooperative work (CSCW) and human-computer interaction (Leonardi et al. 2013) with first papers dating back to 2004 according to Boyd and Ellison (2007). Boyd and Ellison (2007) define Social Network Sites as web-based services with three distinctive characteristics. A user can: (1) set-up a public profile, (2) see a list of user connections and (3) view and traverse social connections of others.

The first paper to coin the term Enterprise Social Networks is from 2006 and called "Enterprise 2.0: The Dawn of Emergent Collaboration" by McAfee (2006). In 2016 Viol and Hess conducted a literature review on publications dealing with Enterprise Social Networks, which I utilise to find relevant literature. They found six meta-topics of Enterprise Social Networks research:

- Implementation
- Motivation
- Usage and Behaviour
- Impact on Organisation
- Success Measurement
- Data and Data Analytics

Usage and Behaviour is concerned with the creation of knowledge (Riemer et al, 2011a; Riemer & Scifleet, 2012), expert search (Richter & Riemer, 2009) and professional versus hedonic uses (Kügler & Smolnik, 2014). I use the research papers for identifying common functionality of Enterprise Social Network software and why organisations are looking to use it. The research topic of **Impact on Organisation** tries to find out how team collaboration and work performance are improved (Alexander, 2015; Kügler et al, 2015b; Suh & Bock, 2015) by using Enterprise Social Networks. I utilise the papers to derive effects on the organisation from the functionality of Enterprise Social Network software. As a starting point for the metric repository the literature on **Data and Data Analytics** along with other literature as described in section 3.4 is utilised. The focus of this literature is to develop analysis approaches for social network data. A common approach is to conceptualise relationships via Social Network Analysis (Behrendt et al. 2014).

	SNS	ESN
User Behaviour	Influenced by site norms	Influenced by organisational policy
Users	Individuals	Employees; use can be optional, encouraged, or mandated
Design	Controlled by a parent corporation, encourages interaction among individual users	Controlled by stakeholders within the organisation, encourages interaction among individual, teams, and other units
Audience	Global or limited to friends	Configured by user or organisational structure
Goals for Use	Hedonic	Professional

Table 1 Key Differences of SNS and ESN adapted from Ellison et al. (2015, p. 107)

Both Social Network Sites and Enterprise Social Networks are social software (Bächle 2006; Boyd and Ellison 2007), where content is created by its users (Richter et al. 2011, p. 90). Therefore the results of research on Social Network Sites is applicable to Enterprise Social Networks, albeit Enterprise Social Networks are used in a professional way (Ellison et al. 2007), while Social Network Sites are used in a hedonic way (Richter and Riemer 2009; Richter et al. 2011). Due to this discrepancy and norms being different inside of an organisation than compared to outside of an organisation, different contexts and environments must be cared for (Richter and Riemer 2009; Richter et al. 2011). That is why Ellison et al. (2015) point out key differences between Enterprise Social Networks and Social Network Sites, which are summarised in Table 1.

2.1.3 Components, Features and Goals

Typical components of social software include webblogs, microblogs, wikis, groups, social bookmarking and instant messaging (Viol and Hess 2016). According to Viol and Hess (2016, p. 352) and Razmerita et al. (2014) the relevant parts for Enterprise Social Networks are wikis for collaboration, document management, social networking and profile pages. Common goals of an Enterprise Social Network include self-presentation and social networking, exchange of information and performing of knowledge work (Riemer et al. 2015). Users have unique profiles and are active in the Enterprise Social Network on a daily basis (Riemer et al. 2015; Ellison and Boyd 2013). Web 2.0 principles such as the collective creation of content, usability and user interaction apply (Viol and Hess 2016).

Leonardi et al. (2013, p. 2) provide a definition for Enterprise Social Networks:

> [An Enterprise Social Network is a] web based platform, that allows workers to (1) communicate messages with specific co-workers or [...] broadcast messages [...]; (2) explicitly indicate or implicitly reveal particular co-workers as communication partners; (3) post, edit, and sort text and files to themselves or others and; (4) view messages, connections, text, and files [...] by anyone else [...] at any time [...].

Distinct features of Enterprise Social Networks are pointed out by Leonardi et al. (2013). The communication between users is usually public and visible. It is straight-forward to publish content in news feeds or groups and published content is persisted and always accessible. Such content is associated with its author and can be discussed by other users (Treem and Leonardi 2012). This results in an inherent instrumental knowledge i.e. "*how to do something*" and meta-knowledge i.e. "*who knows what*" as depicted by Leonardi et al. (2013, p. 4).

2.1.4 Motivation for Use

Enterprise Social Networks provide the users and organisations with a variety of functionality. They change how communication within an organisation takes place by facilitating user participation and interaction (Leonardi et al. 2013).

A common theme is user generated content with emphasis on sharing ideas and knowledge (Boyd and Ellison 2007), which has been proposed by several authors (Mäntymäki and Riemer 2016; Riemer et al. 2015; Zhang et al. 2010). This leads to the exchange of expertise (Steinfield et al. 2009) and ultimately supports the generation of new ideas, brainstorming and problem-solving (Riemer et al. 2012). Based on Kraut et al. (2002), Ellison et al. (2015) and DiMicco et al. (2009) suggest that such exchange of expertise can happen spontaneously in the course of user initiated discussions.

The sharing of knowledge leads to a reduction of knowledge stickiness, which is the act of keeping knowledge to oneself to gain personal benefits (Leonardi and Meyer 2015). As mentioned before, knowledge is publicly available in an Enterprise Social Network and associated with its creators. It leads to open and democratic communication structures (McAfee 2006). This makes the location of knowledge and experts visible and enables teams to communicate across boundaries (Riemer et al. 2015; Jarrahi and Sawyer 2013). This is of special importance to virtual and distributed teams, who otherwise would have trouble identifying experts and locating knowledge (Ellison et al. 2015). In this regard Mäntymäki and Riemer (2016) talk about Enterprise Social Networks fulfilling the information needs of an organisation.

Another aspect is the relationship-building between co-workers. In providing personal information (Ellison et al. 2015) and encouraging relationships between users (Boyd and Ellison 2007), Enterprise Social Networks create ties and enable co-workers to help each other (Mäntymäki and Riemer 2016). DiMicco et al. (2009) talk about *"sense making"* which describes the level of understanding between co-workers. By finding common ground between co-workers, Enterprise Social Networks are helping to increase this level of understanding (Ellison et al. 2007; Jarrahi and Sawyer 2013). This enables employees to integrate into the workforce (Leidner et al. 2010) and build trusting relationships with each other (Ellison et al. 2015).

2.1.5 Effects of Use

The effects of Enterprise Social Network use for the organisation are improved knowledge sharing and transfer between users as well as increased meta-knowledge (Ellison et al. 2015; Leonardi et al. 2013).

Users establish bonding relationships with co-workers and engage in heterogeneous relationships (Ellison et al. 2007; Boyd and Ellison 2007). They develop a sense of corporate citizenship (Steinfield et al. 2009) and thus feel more belonging towards the organisation. This relationship bonding strengthens existing ties and creates new social ties in the organisation (Steinfield et al. 2009), leading to an increased willingness to help and an improved employee performance (Riemer et al. 2015; Kuegler et al. 2015). Dispersed teams in distant locations are able to connect and exchange ideas by utilising Enterprise Social Networks according to Ellison et al. (2015).

Fulk and Yuan (2013) say Enterprise Social Networks are superior compared to traditional knowledge management systems. A 20-25% productivity increase can be gained as mentioned by Mäntymäki and Riemer (2016). A ROI of 365% can be achieved by using Enterprise Social Networks (Mäntymäki and Riemer 2016) and an employee performance increase is stated by Kuegler et al. (2015). This is confirmed by Riemer et al. (2015), who state that active Enterprise Social Network use during project work is positively related to performance. All these benefits of Enterprise Social Networks give it a strategic role in the IT portfolio (Karoui et al. 2015). It should be noted that sustained use is necessary to gain these benefits (Mäntymäki and Riemer 2016) and they are affected by organisational norms, policies and the organisational structure (Ellison et al. 2015; Zammuto et al. 2007).

2.1.6 Groups

Enterprise Social Networks provide various features with regards to groups (Kietzmann et al. 2011). Different types of groups were identified by Muller et al. (2012). Groups can be either set up to be public or private and are usually created for users with a common interest or occupation. These kind of groups usually work together on a shared project or business function and try to achieve the same goal. They discuss specific topics and try to find new ideas related to the project. Other types of groups include technical support groups and recreational groups, devoted to activities unrelated to work.

In general, groups are the centre for collaboration, cooperation and knowledge sharing in an organisation (Riemer et al. 2015). The benefits of Enterprise Social Networks are materialised by using and being active in groups (Bechmann and Lomborg 2012; Nahapiet and Ghoshal 1998). Following this line of thought, I want to measure groups and identify groups, who take a leading role in these activities and in the organisation. Management can use the information on high performing groups to identity domain experts, disseminate knowledge to different groups and establish rewards for top performers. Future research can be conducted on what makes these groups so performant and consequently try to improve other groups.

Groups can have different levels of activity and size, which is relevant for the calculation and interpretation of the metrics later on (Behrendt et al. 2014). The metrics have to be interpreted with the size of the group in mind. The features of a Enterprise Social Network as described by Leonardi et al. (2013) generate network data. This data can be used for an analysis of groups, that is looking to infer conclusions about a group's performance. Typical measures include the time and number of communicated messages and the communication partner.

2.2 Social Capital

Riemer (2005) and Adler and Kwon (2002) performed extensive research on Social Capital. It is based on primary works by Granovetter (1973), Coleman (1990) and Burt (2001). This section summarises their research and puts it into the context of Enterprise Social Networks.

The research on Social Capital is diverse and comprised of multiple theories and perspectives (Adler and Kwon 2002). In general Social Capital describes the value of social relationships in social networks depending on the structure of the network and an actor's position in the network (Riemer 2005, p. 57). Social Capital is an interdisciplinary topic with research in several domains such as social science, economics and psychology (Riemer 2005, pp. 67-70).

Baker (1990, p. 619) defines Social Capital as a "*resource that actors derive from specific social structures and then use to pursue their interests*". A social structure is comprised of social relationships and how they are embedded into the structure of a social network. The value generated from Social Capital can be used to generate business value (Riemer 2005, pp. 68-69). It can be utilised to compare actors based on their social relationships as opposed to comparing their skills and attributes. This is useful to assess situations in which a particular actor performs better than other actors, who have an equal set of skills and attributes (Riemer 2005, pp. 68-69).

Because knowledge work benefits from social relationships, Social Capital is getting more important with the increase of knowledge work (Riemer 2005, pp. 78-79). Especially virtual teams make strong use of software such as Enterprise Social Networks to maintain social relationships and thus Social Capital (Muller et al. 2012). Within such social networks they can cooperate with other actors and access resources, which would otherwise be unavailable to them.

2.2.1 Levels and Perspectives

Actors in a network can be grouped together to form a collective actor (Riemer 2005, pp. 89-94), which can be found on different levels within a network.

As illustrated in Figure 1 on the lowest level are the individual users. Since individuals are atomar actors, they are not considered collective actors. However the next two levels are the group-level and the organisational-level. Both groups and organisations are collective actors, composed of actors from their respective lower level. Groups are composed of different individual actors and their relationships. Organisations are composed of multiple individuals and groups and their relationships.

Steinfield et al. (2009) state that Social Capital can be conceptualised at different levels. All levels can be analysed by using the same theoretical or methodological approach according to Borgatti et al. (1998). They consider "*the substance of Putnam's and Burt's approaches to be separable from the unit of analysis*" (p. 28).

Thus the Social Capital of any collective actor can be analysed according to the two different perspectives proposed by Adler and Kwon (2002): (1) the internal perspective and (2) the external perspective.

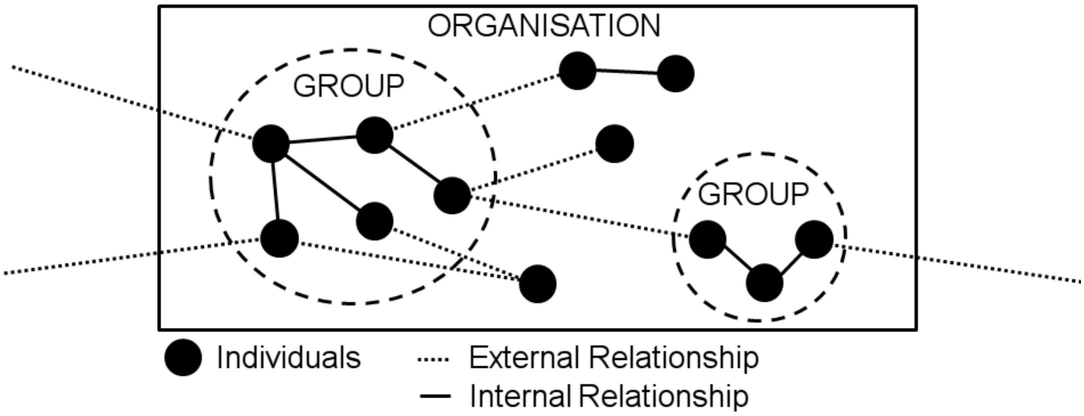


Figure 1 The Different Levels of Social Capital

The (1) internal perspective of Social Capital is concerned with all structures within a collective actor (Riemer 2005, pp. 89-94). Typically the internal perspective is used for analysing the social relationships between group members (Adler and Kwon 2002). Value is generated by cohesion and collaboration within the group (Adler and Kwon 2002). Collaboration is negatively affected by fragmentation and a long distance to other communication partners (Borgatti et al. 1998). As individuals are not considered collective actors, the internal perspective does not apply to them.

Social Capital based on the (2) external perspective is concerned with all interactions between actors outside of the group (Coleman 1990; Sandefur and Laumann 1998). It emphasises the value of direct and indirect social relationships with others. Tymon and Stumpf (2003) claim that it is beneficial for knowledge work as it enables the generation of new ideas. The exposure to new ideas is further supported by heterogeneous relationships to external actors as mentioned by Borgatti et al. (1998).

Basically, the internal perspective is about interactions, that happen within a collective actor and the external perspective is about interactions, that happen outside of the collective actor. As groups are collective actors, both perspectives are suited to analyse groups in a social network.

When analysing Social Capital it is important to take into account what perspective and level you choose as results may differ (Burt 2000; Reagans and Zuckerman 2001). Interdependencies may exist e.g. if many individuals in a group have high Social Capital, it is reasonable to assume, that the group is densely connected internally and therefore also has high Social Capital. There may be conflicting interdependencies e.g. when an individual wants to gain power, it is in the individual's interest to be better connected than others (Blyler and Coff 2003).

2.2.2 Theories about Network Structure

Based on the two perspectives, two theories are proposed as to what constitutes Social Capital: The (1) "*structural holes*" theory was originally proposed by Granovetter (1973) and Burt (1992). It is based on "weak ties" and utilises the external perspective. The (2) "*closure*" theory was originally proposed by Coleman (1988) and Coleman (1990). It is based on "strong ties" and utilises the internal perspective. The connection of the perspectives and the theories is illustrated in Figure 2.

		Perspective	
		Internal	External
Type of Actor	Individual		Structural Holes Weak Ties
	Collective	Closure Strong Ties	Structural Holes Weak Ties

Figure 2 Social Capital Perspectives and Theories

As mentioned before, collective actors can be analysed with both perspectives. Therefore both theories can be applied to collective actors. However, to individual actors only the external perspective and the structural holes theory can be applied.

Structural Holes and Weak Ties

The structural holes theory is based on the external perspective. It focuses on the ego-network of an actor and his structural position. Two assumptions about the actor are made: (1) an actor has a higher number of *diverse*, *weak* ties compared to other actors (Granovetter 1973) and (2) an actor has a *central position* in the network (Burt 1993).

Weak ties are relationships between actors, which are distant and loose e.g. "*some person you know*". The distance to other actors allows them to provide new impulses and non-redundant information (Adler and Kwon 2002). Thus an actor with many weak ties can derive value from novel information and ideas.

The structural holes theory describes a central position in the network as a position, which connects parts of a network, that would otherwise be disconnected (Burt 1993). By connecting different parts of a network, information flows through such a position. An actor in a central position can exert control over the flow of information and is called an information broker (Jansen 2002).

As actors derive value from their positions in the network, they try to get into better positions than their peers (Burt 2001). Therefore the structural holes theory is of competitive nature as actors try to gain an advantageous position over others (Burt 2001).

The type of Social Capital, that the structural holes theory describes, is also called Bridging Social Capital (Putnam 2000, pp. 21f).

Closure and Strong Ties

The closure theory is based on the internal perspective. It focuses on the network as a whole and its internal relationships. It is based on the relevance of strong ties and cohesive network structures (Coleman 1988). Strong ties are created by repeated reciprocal interactions between actors (Steinfield et al. 2009). High cohesion stems from strong ties and close knit relationships within the social network or parts of it (Coleman 1988). This results in a high connectivity between all actors within the network. The network itself derives value, if it is closely connected e.g. enabling effective collaboration (Riemer 2005, p. 108). Thus the closure theory is of cooperative nature as the network as whole gains an advantage.

The type of Social Capital, that the closure theory describes, is also called Bonding Social Capital (Putnam 2000, pp. 21f).

Complementary Theories

Although the theories provide contrasting propositions, they are in fact complementary. This is explained by the fact that the two theories are applied to different perspectives of Social Capital. The structural holes theory is applied to the external perspective, while the strong ties theory is applied to the internal perspective.

Initial work on the complementary character of the two theories was performed by Burt (2000). He states, that if combined correctly, the two theories lead to maximum group performance in an organisation. The idea is that a collective actor should try to maximise Bridging and Bonding Social Capital as shown in Figure 3. The bridging Social Capital is maximised based on the structural holes theory: A high number of weak ties are maintained and used as information resources for gathering non-redundant information and ideas. The bonding Social Capital is maximised based on the closure theory: A cohesive group structure is to be facilitated for effective collaboration and cooperation.

After its maximisation of Social Capital, a group can benefit from several effects. These effects are detailed in the next section.

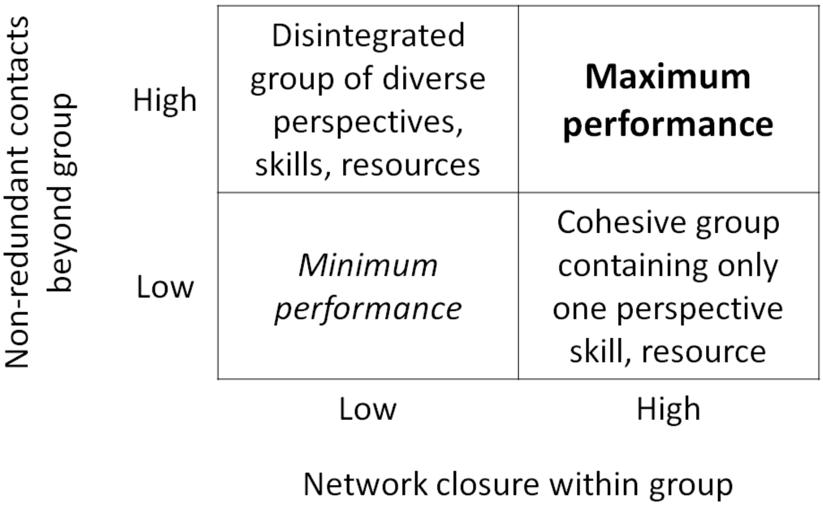


Figure 3 Maximum group performance (Burt 2001, p. 48)

2.2.3 Effects

While the general idea of the two theories is outlined above, I discuss some of the positive and negative effects in detail. Riemer (2005, pp. 116-120), Sandefur and Laumann (1998) and Adler and Kwon (2002) split the effects into categories regarding: (1) information, influence and control and (2) collaboration and cooperation (solidarity).

Positive Effects

(1) Steinfield et al. (2009) links Social Capital to knowledge management. On the one hand the weak ties make novel information and ideas accessible, on the other hand strong ties are suited to debate complex topics and to engage in prolonged discussions (Riemer 2005, p. 117). Scott (2012, pp. 17f) mentions that information is disseminated more quickly via strong ties than weak ties. According to Hansen (1999) weak ties are easy to maintain and therefore a meaningful resource for information.

Based on the structural holes theory, actors in a brokering position can get a lot of influence (Coleman 1990). They are able to access all the information that passes by them and they can decide which information to pass on or withhold (Burt 2000; Riemer 2005, p. 119). This power benefits individuals, who try to use it for their indivdual career progression (Seibert et al. 2001). However such individual influence is not desired by the network as a whole, because it can negatively affect the organisation's Social Capital.

(2) Strong ties and cohesive network structures build norms, reputation and trust (Adler and Kwon 2002; Coleman 1990) based on the principle of reputation and sanction (Burt

2000; Coleman 1990; Riemer 2005, p. 117). This enables effective cooperation (Riemer 2005, p. 118) and the solution of complex problems in an collaborative effort (Krackhardt 1992 and Riemer 2005, p. 118). The norms and trust result in a common ground and shared understanding (Clark and Brennan 1991), which is required to perform knowledge work according to Nahapiet and Ghoshal (1998). Putnam (1995) states that shared norms and trust can lead to societies and subcultures in an organisation, which have positive influence on the organisation's Social Capital. Because users can identify themselves with their respective social network (Watson and Papamarcos 2002), they are motivated and show an increased commitment to their organisation (Scott 2012, pp. 44f). This motivation is linked by Singh et al. (2011) to Social Capital and project success as strongly tied groups sustain solidarity for working together.

Negative Effects

(1) Individuals may exploit their broker position and manipulate information for their own goals (Sandefur and Laumann 1998). This gives them power over the information flow and can result in a disturbed information flow in the organisation.

(2) The establishment of norms and policies can lead to the creation of exclusive clubs or subcultures, which do not share information to the outside (Portes 1998). Such subcultures can oppose the management direction of an organisation in its views and norms according to the organisation culture theory (Hatch 2012). For new employees it can be difficult to enter such closely knit groups, and individual freedom may be restricted by norms and policies (Jansen 2002).

2.3 Enterprise Social Networks and Social Capital

Enterprise Social Networks can provide both Bridging and Bonding Social Capital (Riemer et al. 2015). Ellison et al. (2007) link Enterprise Social Networks to an increased Social Capital in organisations. Ali-Hassan et al. (2015), Riemer et al. (2015) and Kline and Alex-Brown (2013) state the usage of Enterprise Social Networks contributes to the individual's as well as the organisation's Social Capital.

DiMicco et al. (2009) state that Enterprise Social Networks enable users to discover and connect to previously unknown colleagues, facilitating the creation and maintenance of weak ties. Via these weak ties Enterprise Social Networks enable the generation of non-redundant knowledge, ideas and innovation and access to expertise (Steinfield et al. 2009). Based on the provision of weak ties and information access, it can be concluded that Enterprise Social Networks are well suited for accruing Bridging Social Capital (Burke et al. 2011; Ellison et al. 2007).

Enterprise Social Networks are associated with collaboration activities (Richter and Riemer 2013). Fulk and Yuan (2013) and Leonardi et al. (2013) found out that they improve connectivity and interactions among users. Especially distributed or virtual teams benefit from Enterprise Social Networks. The lack of face-to-face communication opportunities allows no spontaneous place-based interactions e.g. in the tea kitchen (Brown and Duguid 2002), which is important for sustaining social relationships (Nardi and Whittaker 2002). Improved communication and collaboration capabilities enable effective problem solving (Riemer et al. 2012), which proliferates Social Capital according to Riemer et al. (2015). Collaboration and effective communication are typical characteristics of Bonding Social Capital. Therefore it can be concluded that intensive use of Enterprise Social Networks leads to strong ties and greater willingness to contribute to the organisation (Ellison et al. 2015). Ultimately, the use of Enterprise Social Networks can lead to higher job performance than no use of Enterprise Social Networks (Riemer et al. 2015).

In an Enterprise Social Network social ties can be established by adding friends, writing messages or participating in groups and threads. These ties help to create and maintain Social Capital in the network (Steinfield et al. 2009). Even in distributed organisations, it is possible to have reciprocal social connections i.e. strong ties. Users are able to add acquaintances and "friends of friends" i.e. new contacts and weak ties.

It should be noted that there are Enterprise Social Networks, in which information and relationships are public, so information brokers as mentioned in the Structural Holes theory have less influence. In this sense Enterprise Social Networks can reduce the need and opportunity for bridging in an organisation (Riemer et al. 2015) as the public knowledge can be obtained directly.

As it pushes competitiveness among employees (Portes 1998), information brokerage has negative association with organisations. Riemer et al. (2015) state that middle managers are important for passing along information and might be obsolete with the public information in Enterprise Social Networks. Thus middle managers may disapprove the introduction of Enterprise Social Networks as information hoarding is less feasible with public information (Riemer et al. 2015).

In any case Enterprise Social Networks decouple information flow from personal communication channels such as email or phone to public information channels in the network (Riemer et al. 2015). Information is freely accessible and collaboration can be made more effective.

I want to present two examples to demonstrate Bridging and Bonding Social Capital in Enterprise Social Networks.

(1) Bridging Social Capital: An example from Enterprise Social Networks would be a user, that has a big contact list with people from all units of the organisation. On the news feed the user receives frequent information from all the units and if the user requires particular domain knowledge, it can be utilised via one of its relationships.

(2) Bonding Social Capital: An example from Enterprise Social Networks would be group-members engaging regularly in discussion and working collaboratively. The frequency of their interactions leads to strong ties, which in turn enable their effective collaboration.

2.4 Social Network Analysis

After having discussed how Enterprise Social Networks constitute for Social Capital in theory, my goal is to analyse Enterprise Social Network data and measure Social Capital for particular networks. Desired findings of such an analysis are how Social Capital impacts the individuals, groups and organisations within the network (Wasserman and Faust 1994, p. 10).

The research of Social Network Analysis is emergent interdisciplinary field (Stieglitz et al. 2014). In early research manual, qualitative studies with questionnaires (Hacker et al. 2016) were conducted. Manual collection of Social Network data is time-consuming and costly (Fischbach et al. 2008). In the last five years, with the adoption of Enterprise Social Network platforms, their large amount of data can be leveraged for analysis.

Stieglitz et al. (2014) developed a Social Network Analytics framework, which provides an overview of research domains, approaches and methods in social networks. It distinguishes three main analysis methods: (1) text mining, (2) Social Network Analysis and (3) trend analysis. My focus is (2) Social Network Analysis which I outline in the following.

Social Capital is intangible (Riemer et al. 2015) and cannot be quantified or measured directly (Adler and Kwon 2002). Therefore Social Network Analysis operationalises the network as a social network graph that can be measured. Statements about Social Capital are inferred from the social structures in the graph such as a user's position or relationships. When interpreting the measures, it should be noted that datasets are biased towards origin and context. Thus the context of the data must be considered to derive meaningful conclusions (Stieglitz et al. 2014).

2.4.1 Explanation of SNA

The Social Network Analytics approach models relationships between persons, organisations and groups based on a graph theoretic model[4] (Scott 2012). It considers knowledge embedded in organisational structure (Brown and Duguid 2001) with a focus on relationships (Hansen 1999). Linkages between employees and their social interactions are the subject of analysis (Allen et al. 2007). Therefore it is well suited to operationalise Social Capital. Typical graph metrics and approaches can be used for the analysis. Essential approaches and benefits are discussed in this section. The collected metrics are applied to Enterprise Social Networks and my dataset in section 4.

To model the Enterprise Social Network as a social network graph, the actors are represented by graph nodes and social relationships are represented by directed edges (Newman 2010, p. 109). This is illustrated in Figure 4. Since actors can be individuals, groups or organisations, a node can represent either of them (Scott 2012, p. 9). An edge can have different meanings: while it represents a social relationship, the type of the relationships is undetermined. For example, it can represent a post, a like or any kind of interaction (Newman 2010, p. 37) which is possible in the social network. The intensity or strength of the relationship is described by the edge weight (Scott 2012, pp. 30f)

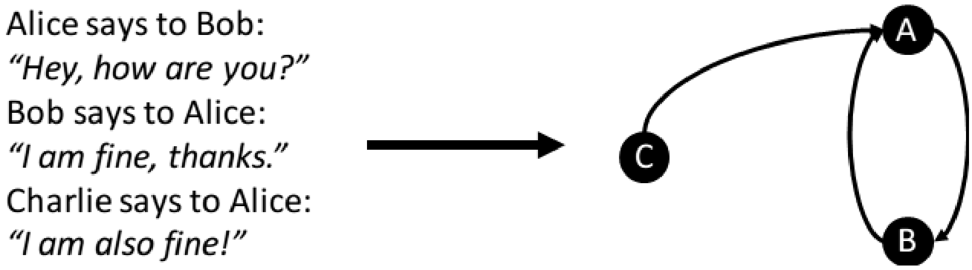


Figure 4 Social Network Graph

In Figure 4 Alice and Bob are modeled as the nodes A and B. They exchange two messages with each other, that are modeled as edges between the two nodes. Charlie is the node C and sends Alice a message. Since Alice does not respond, there is only one directed edge from Charlie to Alice, but no edge back.

When modeling a network as a graph, Wasserman and Faust (1994, p. 20) describe a social network as "*a finite set or sets of actors and the relations defined on them*". Key concept according to Wasserman and Faust (1994) is the focus on social relationships

[4] Graph theory is the mathematical study of nodes and edges, whereas edges connect the nodes. The edges can be directed or undirected i.e. having a direction. They can also have weight, describing the strength of the edge. Many algorithms and mathematical models exist to describe graphs and calculate their attributes, some of them can be useful for Social Network Analysis.

and their inherent value. They (p. 4) note the following four principles: (1) actors and actions are interdependent; (2) relationalships between actors are of value by "transporting" resources e.g. information; (3) network structure can support or constrain individual actions; (4) network models conceptualise social structure.

2.4.2 Benefits

Social Network Analysis has practical relevance as it helps decision-making with human resource and knowledge management (Stieglitz et al. 2014). It can be used to detect influential (according to Social Capital theory) members in an Enterpise Social Network (Cross and Prusak 2002). Based on the graph metrics, high performing users and groups can be identified. Social Network Analysis can detect how information is diffused through the network and how the information flow is shaped by the network (Scott 2012, pp. 42-44). I want to use it to find out where new ideas and information are exposed in groups and which groups are high-performing.

2.4.3 Applying Social Network Analysis

After the social network graph is constructed, graph theoretic metrics can be applied. They are distinguished into two categories: (1) egocentric measures and (2) global measures.

(1) Ego-centric measures are based on a focal individual i.e. center of analysis puts the indvidual in the center of the network. It is concerned with the neighborhood of the individual. This means they are calculated from all direct and indirect relationships and neighors this individual has. A direct relationship means the indivdual is connected personally to the other end, whereas an indirect relationship describes a situation where you know someone who knows the other. Such indirect relationships can be of different lengths, meaning that they can be connected via a path of neighbours (cf. Xing or LinkedIn social network).

(2) Global measures are concerned with the properties of the network as a whole. The measures are calculated from the set of all the actors and relationships in the network. They are not simple aggregates but structural features of the network. An example measure is how densely connected the network is as a whole i.e. referring to Social Capital's cohesion.

For my Metric Repository in chapter 4 there is one major difference to the notion of ego-centric measures and global measures: I define the two categories based on whether the interpretation is concerned with a focal individual or a global network, whereas the literature classifies the two categories based on whether the calculation is local or global. I changed this, because in this case it is easier to understand and compare the metrics' interpretations. So it is straight-forward for readers of this thesis to understand the results.

It should be noted that a minimum of activity and group size is required for an analysis, otherwise false results occur (Riemer et al. 2015). To identify key employees, it is mandatory that the employees actively use the Enterprise Social Network (Behrendt et al. 2014). The interpretation of the metrics' values varies on the context and size of the social network and its underlying data.

2.5 Summary of Background

Enterprise Social Networks are increasingly adopted by organisations. Employees of an organisation can use the platform to create messages, present themselves and communicate with others.

Motivation for the use of Enterprise Social Networks comes from the generation of new ideas and the sharing of knowledge between employees. Social platforms prove as a valuable tool for communication in teams, especially if these are virtual or dispersed.

The communication and organisation of employees in groups, facilitates relationships between co-workers. What value these relationships provide for the organisation is discussed in the Social Capital theory literature.

The Social Capital research considers several theories in the context of social networks. In these theories the employees of an organisation are described as actors and groups are called collective actors. A common approach is to take an internal and external perspective on the relationships of an actor. The structural holes theory takes the external perspective and examines the value of relationships between different individual actors outside of a group. This is called Bridging Social Capital. Contrary, the closure theory takes the internal perspective and discusses the value of relationships within a collective actor. This is called Bonding Social Capital.

To measure Bridging and Bonding Social Capital in Enterprise Social Networks, it can be operationalised via Social Network Analysis. Social Network Analysis models the social structure of a network as a graph theoretic model and allows the application and calculation of common network metrics. Based on these metrics I want to detect influential and high-performing groups in Enterprise Social Networks. This is described in the following chapters.

3 Research Methodology

I want to measure group metrics in the context of Enterprise Social Networks and interpret it against the theoretical underpinnings. For this cause Swoop provided me with a Yammer data set.

In the following I describe the organisation Swoop and their needs and I describe the Enterprise Social Network software Yammer, which is the source of the data. Specifically I look into the features of Yammer and how the structure of the data set looks like. In a next step I explain my research approach for creating the metric repository and provide an overview on how I propose to measure the identified metrics for groups.

3.1 Swoop

Swoop is an Australian organisation based in Sydney, founded by Cai and Mariannae Kjaer and Dr Laurence Lock Lee in the year 2014. With a team of ten people they are working to help people become better collaborators[5].

Swoop provides a unique platform for Social Network Analytics that gives individuals, teams, communities and executives insights into how they work together. It is the leading Social Network Analysis platform and presents a diverse set of valuable social network visualisation tools[6].

Figure 5
Swoop Logo

As an organisation Swoop passionately believes in the power of collaboration and people networks to get work done. By analysing data from a variety of sources they provide employees and management with insights to make informed and evidence-based decisions about collaboration and the health of social networks[6].

Their product is the result of more than 10 years of consulting experience in mapping organisational networks. Based on more than 100 projects the founders have identified the most valuable metrics that help organisations drive collaborative business performance[6].

The targeted market is global and attracts a rapidly growing customer base in the US, Europe and Australia. Global top customers of Swoop include but are not limited to[6]: Yum! Brands, the owner of KFC, which is linking disparate employees across their many

5 http://www.swoopanalytics.com/index.php/about/ (accessed 2017-02-07).
6 http://www.swoopanalytics.com/ (accessed 2017-02-07).

outlets by using Swoop's platform. Telstra, with more than 40,000 employees, is among the top five companies listed on the Australian Stock Exchange.

To gain a better understanding of the networks usage and to identify highly performing users, Swoop's platform evaluates data in five levels[6]:

(1) Enterprise – uses data from the entire network,
(2) Groups – uses data from groups,
(3) Topics – uses data from one or multiple topics,
(4) Business Unit – uses data from a business unit and
(5) Personal – uses data from an individual user.

Based on the used data, Swoop calculates different metrics for each level and the individuals from the particular level. The metrics are visualised in an online frontend for users to see. The visualisation includes the number of interactions that have taken place on the given level, e.g. the number of posts, replies or likes, and the network activity, e.g. activity per user, response rate, or the ratio of public/private messages. Typical visualisation of these metrics includes tables and graphs. Except for the personal level, the platform identifies key players.

Currently, Swoop wants to add metrics on the enterpise and group level. Swoop's and my goal is to identify key groups based on metrics from the latest scientifc publications. These key groups and their metrics are to be visualised in a web frontend using Swoop's technology stack. The data for the analysis is sourced from the Enterprise Social Network Yammer, which is used by Swoop's customers[7].

3.2 Yammer

Yammer is an enterprise social networking service used within organisations as part of Microsoft's Office365 suite. The company Yammer was founded in 2008 by David Snacks and acquired by Microsoft in 2012. More than 70% of Fortune 500 corporations employ the Yammer platform in their organisation. According to Gartner, Microsoft is the leading vendor of 2015 for social sotware in the workplace with this product[8]. The platform is accessed via webinterface and the domain of the user's email is used to bind one to the community. For example the email "test@uni-sydney.edu" would be assigned to the "uni-sydney.edu" community as the communities are based on domains[9].

[7] This was discussed in an internal meeting with Swoop on 2016-09-14.
[8] https://blogs.office.com/2015/10/28/gartner-recognizes-microsoft-as-a-leader-in-the-2015-magic-quadrant-for-social-software-in-the-workplace-for-seven-years-running/ (accessed 2016-12-16).
[9] https://products.office.com/en-us/yammer/yammer-overview (accessed 2017-01-04).

Yammer is used to connect employees across an organisation. The focus is on the discussion of ideas, sharing of updates, and crowdsourcing of answers from coworkers around the globe. "*Yammer gives your team a faster, smarter way to connect and collaborate across your company*"[9] according to Microsoft's product page. In general, the goal is to share knowledge across the organisation while integrating well with other Office365 products.

The functionality of Yammer includes a public timeline in the domain community, the opportunity to create public and private groups as well as to create threads and initiate discussions in domains or groups. Users reply to threads, react on likes and mentions and can vote in polls. Private instant messaging and sharing documents with other users is possible[9].

3.3 Data Structure

All user actions are persisted in the Yammer backend. Swoop retrieves this data via Yammer's application programming interface and stores it in their own database. The part of the database, which is relevant for me is depicted in Figure 6.

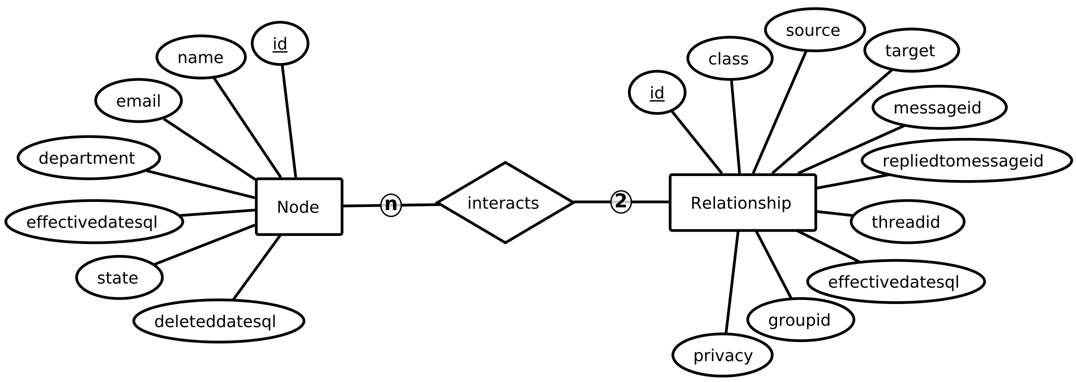


Figure 6 Swoop's Data Structure

The *Node* entity represents a user in the network and it has 7 relevant attributes, that I use in my analysis. The attribute `id` is a numerical primary key, the attributes `name`, `email` and `department` are of type text and describe the user. `effectivedatesql` is the creation date of the user account. `state` is the current state of the account and can have on of the following values: `soft_delete`, `active`, `external`. Based on the state the user can either be an active member of the community, an external user, that has access to the network[10] or a user with a deleted account. If the account is deleted, the attribute `deleteddatesql` contains the date of the deletion.

Users in the social network can interact via five classes of communication with each other. All interactions are described by the entity *Relationship*. Each interaction is given a unique primary key called `id`.

All interactions are part of a particular thread, identified by the attribute `threadid`. The class of the interaction is stored in the attribute `class` and can have one of five values: `Reply`, `Like`, `Post`, `Notification` or `Mention`. A `Post` is the first message of a thread and a `Reply` is a response to a `Post` i.e. the second, third or later message in a thread. `Notification` and `Mention` interactions are always part of a `Post` or `Reply`. They are used to notify other users of the particular message. A `Like` can be given to any message as a form of acknowledgement without actually creating a textual message. While an interaction is identified by the `id` attribute, the message is identified by `messageid`. This allows it, to attach `Notification` and `Mention` interactions to a message, but store them as separate interactions in the database. It should be noted, that while messages can contain multiple interactions of class `Notification` and `Mention`, they can only have one of either `Reply` or `Post` interaction.

The `source` attribute identifies the source `Node` entity, which authored an interaction. The `target` attribute identifies the target `Node` entity, to which the interaction is directed.

`effctivedatesql` is the creation date of the interaction and `privacy` describes the publicy of an interaction, which can be of value `Public` or `Private`. The `groupid` attribute identifies the group in which the interaction took place and thus is important for the group analysis.

The different classes used in Swoop's date structure can be found in the literature. Viol and Hess (2016) distinguish classes of relationships, e.g. posts and likes, into information seeking, information sending and info receiving types of content.

Swoop provides anonymised data, that means all attributes related to message content or personal user information are left blank. For reasons of relevancy and visibility in the Figure 6 the following attributes were left out:

- for Relationship: isreciprocal, relationshipid, updatetime
- for Node: image, effectivedate, updatetime, deleteddate, departmentid

[10] cf. section 3.2. Yammer communities are domain-based. All users outside of the domain, are external users.

It is possible to discover distinct behavioural dimensions without having to evaluate message content (Viol and Hess 2016). While the metrics in the metric repository in section 4 take into account metrics based on message content, they cannot be considered in the prototype in section 5 due to lack of data.

3.4 Research Approach

My aim is to create a reusable metric repository containing all metrics, that have been proposed in the literature until the end of 2016. This metric repository includes the interpretation of each metric and its calculation schema. Therefore, it can be used as a resource for developing software systems that analyse the Social Capital of Enterprise Social Networks. Based on the metric repository I build a prototypical web platform, that presents group metrics from Enterprise Social Network data.

Riemer and colleagues published articles on this topic, which I use as a starting point for the collection of metrics. I look at their publications and the publications of their co-authors with the latest papers being released in 2016. Starting from these papers I perform a backwards search to identify further literature on the topic of Enterprise Social Networks and their metrics. The main sources that I use for the metric repository, are Smith et al. (2009), Freeman (1977), Hacker et al. (2015), Viol and Hess (2016), Berger et al. (2014), Wasserman and Faust (1994), Scott (2012), Newman (2010) and Angeletou et al. (2011). If required, additional background information is retrieved from secondary sources.

Having identified the metrics, I discuss their possible interpretations against the backdrop of Social Capital theory. For the theoretical background on Social Capital, I started with Riemer (2005), Adler and Kwon (2002), Burt (2001), Coleman (1988) and Granovetter (1973) as they are main works on the Social Capital theory.

The software prototype puts an emphasis on providing automated calculations of the metrics. Thus, I implement a backend system, which can calculate and store all metrics without user interaction. The analysis results are published via a web service. A web frontend consumes this web service and provides up to date visualisation of the data. The visualisation design is based on the practical guidelines described in the IBCS Standard[11]. It emerged from the academic works of Tufte (2001), Minto (2003), Few (2006), Shneiderman (1996) and Brinton (1914) and was edited by Hichert and Gerths (2011).

[11] http://www.ibcs-a.org/standards (accessed 2017-02-07).

3.5 From Metrics to Group Metrics

The metric repository is a general reusable metric repository, that is not specifically adapted to group metrics. It contains ego-centric metrics, that are calculated for individual actors and global metrics, that are calculated for the entire network. My goal is to apply these general metrics to groups, while I utilise both the ego-centric metrics and global metrics. By using ego-centric and global metric, I measure the internal perspective (Structural Holes theory) and the external perspective (Closure theory) as described by Borgatti et al. (1998) and Burt (2001).

The global metrics are well suited to measure a group's internal perspective of Social Capital. For the analysis of a particular group, I filter out all interactions that do not take place within said group. The result is, that only the internal communication of the group is left. I use this communication to calculate the global metrics for the particular group. From the results of the metric calculation, I can infer conclusions about the internal perspective of Social Capital.

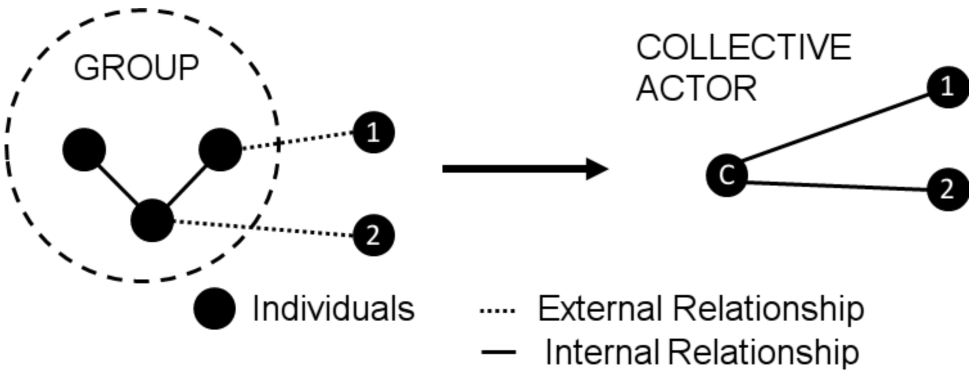


Figure 7 Modeling Groups as Collective Actors

I use the ego-centric metrics to measure a group's external perspective of Social Capital. Since the ego-centric metrics are not adapted to groups, they cannot be calculated for groups without preparation. To overcome this issue, I propose to model the groups as collective actors based on the Social Capital theory in section 2.2. According to Adler and Kwon (2002) individuals can be grouped together to form a collective actor. For collective actors the same set of measures can be used as for individuals (Borgatti et al. 1998).

My approach is illustrated in Figure 7. I take all users within a particular group and merge them to a new node, which is the collective actor representing the group.

All incoming and outgoing interactions from the original users are redirected to the new collective actor. All internal interactions of the particular group are filtered out, so only external interactions are left. I use the external interactions of the collective actor to calculate the ego-centric metrics for the group. From the results of the metric calculation, I infer conclusions about the external perspective of Social Capital.

With my approach, both the internal and external perspective of Social Capital can be measured. Based on Burt's (2001) complementary theory, both perspectives are taken into account to determine the performance of groups.

4 Metric Repository

Different authors have developed metrics to identify key users and classify them into roles. I want to utilise those metrics to analyse groups. Therefore I identify the metrics from the literature and compile a repository of metrics. Due to the number of 62 metrics, I structure and filter them based on their attributes. This leaves 41 metrics for a detailed discussion. In the following section I explain their main attributes, how to calculate and how to interpret them.

4.1 Repository Structure

Each metric is given a **name**, which is human-readable and identifies the metric. The metric's **numerical id** allows me to specify ranges of metrics and it can be reused in algorithmic operationalisation later on.

All metrics were sourced from the **literature** and therefore all authors are referenced, who proposed a specific metric. It is possible that multiple authors proposed the same metric. If less than three authors proposed a metric, it is not discussed in detail. Instead it is listed with its attributes in the appendix.

Because the metrics differ with regards to their origin and calculation, they are distinguished into two **types**: (1) Graph metrics and (2) Enterprise Social Network metrics.

(1) Graph metrics are derived from graph theoretic measures,
(2) Enterprise Social Network metrics are values given from my data set.

The graph metrics are further divided into ego-centric and global scope. The Enterprise Social Network metrics are divided into single metrics and combined metrics. Single metrics do not require a calcuation i.e. they can be read from a column in the database, directly. The combined metrics require a combination of different columns in the database.

The **scope** specifies whether a metric is relevant for an individual node and its neighbourhood (ego-centric) or if a metric is relevant for the entire network or group as a whole (global). It is a term from the domain of graph theory, but can also be applied to the Enterprise Social Network measures. Besides ego-centric and global, a metric can also be relevant for both. If both scopes are relevant, the calculation and interpretation of a metric can differ depending on the scope. Metrics with ego-centric scope measure the external perspective of Social Capital and metrics with global scope measure the internal perspective of Social Capital.

To clear up towards the readers, what value each metric calculates, a short textual **description** is provided. The description refers to the technical, theoretic perspective of the metric and already hints verbally on how to calculate the metric.

For the implementation of the metrics in software, it is necessary to provide a formal definition on how to **calculate** each metric. The calculation schema is relevant for the Social Network Analysis metrics and Enterprise Social Network metrics. With regards to the Enterprise Social Network metrics, the calculation schema refers to the correct column or query (in case of a count, sum or average) of the data set.

The **interpretation** of a metric is essential, because it tells you how to understand the value of the metric. It answers the question what a low or high value of each metrics means. For example, a high number of posts can mean that a user is very active in the social network, while a low value indicates the opposite. Different authors provide their interpretations on metrics and try to classify users into user roles. These user roles and the metrics' interpretations are discussed against the backdrop of the Social Capital theory.

As Social Network Analysis metrics and Enterprise Social Network metrics can refer to the same value and conceptual dimension, there can be **overlaps** in the metric repository. For example the Social Network Analysis *degree* metric of a node is equal to the Enterprise Social Network metric *count of distinct interaction partners*. It is relevant to identify overlaps and acknowledge that the value and dimension may be the same, so there is no confusion in case values and interpretations are duplicate. To keep the later evaluation and visualisation implementation of the results simple, a single metric should be chosen for displaying a specific dimension.

The different attributes structure the repository of metrics. This gives an overview of the metrics and makes it the reusable for further research. For this thesis, the different attributes influence how to calculate the metrics and what metrics to visualise.

4.2 Graph Theoretic Metrics

To utilise metrics based on a graph theoretic model, I model the data as a graph. All interactions in the data have a source user and a target user. Thus I define the Social Network as a directed graph G, which contains all users and their respective interactions:

$$G = (V, E).$$

V is the set of all vertices, which represents the set of users in the Social Network. E is the set of ordered pairs called edges, which represents the set of directed interactions of the users. If multiple interactions between users occurred, this is represented by setting the weight of the edge between the two users. As the relevant features are the positions and edges between the nodes, the type of interaction (post or like) is not modelled in the graph.

The adjacency matrix $A = [a_{ij}]$ is a matrix, that is the cartesian product of all the nodes in the graph. Its entries are defined such as the value $a_{i,j} = c$ denotes an edge from the node v_i to the node v_j. The value c denotes the weight of an edge or 0 if there is no such edge

$$a_{ij} = \begin{cases} c_{ij} & (i, j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

For the calculation it is important to note, that the edge weight is interpreted as weight strength and not weight costs. This means, that the weights need to be inverted for some implementations[12]. Otherwise metrics such as closeness will provide inaccurate values.

To structure the metric repository, the following measures are divided up into two sections: the ego-centric metrics and the global metrics.

4.2.1 Ego-centric Metrics

In the following an in-depth look at ego-centric metrics is conducted. Because it is relevant for the implementation, an emphasis is put on the calculation. The interpretation of the values is important for the later visualisation and frontend. Therefore another emphasis is put on the interpretation of the metrics with regards to Social Capital.

Degree Centrality

The *degree* or *degree centrality* measures the number of edges connected or adjacent to a vertex $v \mid v \in V$ (Newman 2010). In directed graphs such as mine, it can be split up into in-degree and out-degree, reflecting the incoming edges to a node and the outgoing edges from a node (Newman 2010). Since the degree measure counts the edges for a specific node, the scope of the measure is ego-centric. It shows how strongly connected a node is in terms of relationships with other nodes. This metric is proposed by Smith et al. (2009), Hacker et al. (2015), Viol and Hess (2016), Angeletou et al. (2011) and Berger et al. (2014).

[12] This is relevant for my implementation as it is based on igraph. igraph models the weight as costs from A to B, instead of strength between A and B.

The in-degree of a node $d_{in}(v_i) \mid v_i \in V$ is equal to the number of edges e_k in the form of

$$e_k = (v_j, v_i) \qquad \text{for all } e_k \in E \text{ and } v_j \in N.$$

The out-degree of a node $d_{out}(v_i) \mid v_i \in V$ is equal to the number of edges e_k in the form of

$$e_k = (v_i, v_j) \qquad \text{for all } e_k \in E \text{ and } v_j \in N.$$

The degrees can be calculated by using the adjacency matrix:

$$d_{in}(v_i) = \sum_{v_j \in V} a_{j,i} \qquad d_{out}(v_i) = \sum_{v_j \in V} a_{i,j}.$$

A problem with the degree is that its interpretation depends on the size of the network g. To compare the degree of differently sized networks, Wasserman and Faust (1994, p. 178) propose the following standardisation through dividing by the maximum possible degree – which is the network size g minus one:

$$d'_{in}(v_i) = \frac{d_{in}(v_i)}{g-1} \qquad d'_{out}(v_i) = \frac{d_{out}(v_i)}{g-1}.$$

With regards to Social Networks Wasserman and Faust (1994, p. 126) define the in-degree as a measure of popularity (incoming messages) and the out-degree as a measure of expansiveness (outgoing messages). Wasserman and Faust (1994, p. 179) state that a high degree centrality is recognised as a major channel of information. Newman (2010, p. 169) adds that a user with a high degree centrality and thus a high number of connections to others may have more influence than users with a lower degree centrality. Therefore a high degree centrality is an indicator for key users. This is reinforced by Berger et al. (2014), who claim that a high in-degree is distinctive of key users.

According to Angeletou et al. (2011) a low in-degree indicates an *elitist* user. Such a user communicates with only a small group of other users, but has strong reciprocal interactions with those users. A high in-degree indicates a *popular initiator* and *participant* kind of user. This type of user contributes with a high intensity, persistence and reciprocity to many other users (Angeletou et al. 2011). *Elitists* and *popular users* drive the discussion and increase the activity of a community, making information available and interactions feasible (Angeletou et al. 2011).

Smith et al. (2009) correlate a high degree centrality with an *answer person* and *discussion person*, seeking to actively engage in other people's threads. They participate in discussions of considerable length. He describes those kind of people as influencers, which aligns with other literature. Contrary to the influence indication, the degree metric is not a direct indicator of a user's performance according to Riemer et al. (2015). Thus an influential user does not automatically make a productive employee.

If multiple users exhibit a high degree centrality, it leads to a dense and cohesive network. A cohesive network structure with redundant relationships, also called "closure" (Riemer 2005, p. pp. 107f), leads to the creation of Social Capital according to Coleman (1990). Characteristics of such a network include a collective mindset and effective norms, which results in Social Capital (Riemer 2005, p. 107). Cohesive networks and effective norms are required for cooperation and trust in networks (Riemer 2005, p. 108), which ultimately leads to superior performance.

Closeness Centrality

The *closeness centrality* measures the average shortest path, also known as the geodesic distance, through a network between two vertices (Newman 2010, p. 181). It was first published by Sabidussi (1966). Since the closeness is the opposite of the distance, the value of the average shortest path is inverted to get the value for the closeness centrality. A higher value for the closeness centrality means that a vertex is closer to all other vertices (Scott 2012, pp. 33-34) and quicker to interact with all others vertices (Wasserman and Faust 1994, p.181).

Because the closeness centrality measures the distance from one specific node to all others, it is of ego-centric scope. This metric is proposed by Smith et al. (2009), Hacker et al. (2015), Viol and Hess (2016) and Berger et al. (2014).

The calculation is based on Wasserman and Faust (1994, pp. 184f). To calculate the closeness centrality, the number of edges between the two nodes $v_i \in V$ and $v_j \in V$ for the shortest path is defined as the $distance(v_i, v_j)$ – this equates to the shortest path length or the geodesic distance.

To get the closeness centrality c, the sum of all distances is calculated and inverted:

$$c(v_i) = \left[\sum_{v_j \in V} distance(v_i, v_j) \right]^{-1} \quad \text{, where } v_j \neq v_i.$$

For comparison of networks with different sizes, Wasserman and Faust (1994, p.185) as well as Freeman (1979, p. 225) pick up the suggestion by Beauchamp (1965) to standardise the metric through multiplying with the network size g minus one:

$$c^{'}(v_i) = (g-1) * \left[\sum_{v_j \in V} distance(v_i, v_j) \right]^{-1} \quad \text{, where } v_j \neq v_i.$$

Because the formula yields an infinite value for disconnected nodes, it is problematic. While Newman (2010, p. 184) suggests to use the inverse distance instead of the distance, he acknowledges that this is rarely used in practice. Instead the formula variant from the igraph package is used, which states that "*if there is no path between vertex v_i* and v_j *then the total number of vertices is used in the formula instead of the path length*"[13].

Therefore I define a new distance function as:

$$distance'(v_i, v_j) = \begin{cases} distance(v_i, v_j) & \text{, if there is a path} \\ g & \text{, otherwise.} \end{cases}$$

The final formula looks like this:

$$c'(v_i) = (g-1) * \left[\sum_{v_j \in V} distance'(v_i, v_j) \right]^{-1} \quad \text{, where } v_j \neq v_i.$$

The value can be interpreted in the context of Social Networks. Newman (2010, pp. 183-184) mentions that a user with high closeness centrality is able to quickly respond and interact with other users due to the short distance. Such a user can efficiently disseminate information through the network because of the short communication paths to others. This argument is confirmed by Berger et al. (2014), who state that users with high closeness centrality are able to spread information easily. Similar to the interpretation of degree centrality, Berger et al. (2014) claim that a high closeness centrality is related to being a key user. Key users are knowledge hubs, meaning they contribute and help other users to solve their daily problems. They are able to diffuse innovative ideas quickly to other people.

Smith et al. (2009) relate high closeness centrality to users who regularly spawn new discussions and ideas as well as take part in other users' threads. Contrary, a low closeness centrality is related to people who tend to reply in other people threads only, but do not initiate discussions on their own. The notion of engagement is introduced by Hacker et al. (2015). On the one hand a high closeness centrality indicates high levels of continuous engagement by a user in the Enterprise Social Network, but on the other hand such a user is not very focused i.e. the user talks about multiple topics compared to being an expert in one topic.

Hacker et al. (2015) draw connections to Viégas (2004), who researched that the closeness centrality can be related to the frequency of posts. A high closeness centrality and degree centrality indicate a high post count, which is discussed in the metric *messages created*. Furthermore Hacker et al. (2015) links to Holtzblatt et al. (2013) and their results on

[13] http://igraph.org/r/doc/closeness.html (accessed 2016-10-25).

valuable themes of social platform experience. They claim a high engagement supports collaboration and facilitates cooperation with staff in other locations. It also strengthens social connection, expanding a user's network and tracking other people's activities. Since a high closeness centrality implies that you can easily communicate with all other users, both of these statements are reasonable to make.

Viol and Hess (2016) pick up the interpretation of continuous engagement in an Enterprise Social Network. A high closeness centrality indicates that a user is well connected within the network. They are always online and active and therefore can initiate and take part in multiple discussions. This means that they are not focused on one topic, but rather dispersed across a lot of discussions and threads (Viol and Hess 2016), which fits to the interpretation in the other literature.

The interpretation is similar to the degree centrality. If multiple users exhibit a high closeness centrality, it leads to a dense and cohesive network. The short distance between all users results in strong ties between the users. Strong ties are a reason for Social Capital and the formation of effective norms and trust (Riemer 2005, p. 108). According to Coleman (1990) effective norms and trust among the users allow for successful collaboration.

The interpretation that information can quickly be disseminated, fits to the effect of Social Capital described by Nahapiet and Ghoshal (1998). They argue that "*Social Capital constitutes a valuable source of information benefits*" (p. 252). It manifests itself in the distribution of information, making information readily available. Because of the established trust through strong ties, specifically the distribution of complex or sensitive information is positively influenced by Social Capital, as noted by Riemer (2005, p. 117) and Koka and Prescott (2002, p. 801). Whereas the dissemination of arbitrary and small information is more positively influenced by weak ties and thus a low closeness centrality.

Betweenness Centrality

The *betweenness centrality* measures the likeliness that a node is an intermediary between any other two nodes in the network (Wasserman and Faust 1994, p. 189). A node with a high betweenness centrality may have high influence in the network as a lot of information is passed by this node (Newman 2010, p. 186). A high betweenness centrality is achieved, when a node fills a high number of structural holes and it is in a brokering position (Scott 2012).

Because the betweenness centrality measures the betweenness for a particular node in the network, it is of ego-centric scope. The metric is proposed by Smith et al. (2009), Hacker et al. (2015), Viol and Hess (2016) and Berger et al. (2014) and discussed in Wasserman and Faust (1994).

Mathematically a node is in this position if it lies on the geodesic path of two other nodes. According to Freeman (1977) the betweenness centrality of a node v is defined as $b(v_i)$ and is calculated by the probability of the node v being on any geodesic path between two other nodes v_j and v_k with $i \neq j \neq k$. The number of geodesic paths between the nodes v_j and v_k is defined as $geop_{j,k}$. It is assumed that all of these paths are equally probable to be chosen for a communication action. Thus, the probability for a communication action using a particular path is $1/geop_{j,k}$. Furthermore, I define $geop_{j,k}(v_i)$ as the number of geodesic paths that contain the node v_i. This results in the following formula describing the probability that the node v_i falls on a randomly selected geodesic path between v_j and v_k:

$$p_{geop}(v_i) = \frac{1}{geop_{j,k}} * geop_{j,k}(v_i),$$

which is shortened to:

$$p_{geop}(v_i) = \frac{geop_{j,k}(v_i)}{geop_{j,k}}.$$

For getting to the betweenness centrality $b(v_i)$, the sum of the above probability over all unordered pairs of nodes not including v_i is calculated:

$$b(v_i) = \sum_{j \neq k} p_{geop}(v_i),$$

or written as:

$$b(v_i) = \sum_{j \neq k} \frac{geop_{j,k}(v_i)}{geop_{j,k}}.$$

Wasserman and Faust (1994) suggest a normalisation of Freeman's formula, similar to degree- and closeness centrality. In a network of size g, the maximum value for the betweenness centrality is $(g-1)(g-2)/2$, in case the node v_i lies on each of the geodesic paths. Therefore I normalise the formula to a final version:

$$b'(v_i) = \frac{b(v_i)}{(g-1)(g-2)/2}.$$

Since a high betweenness centrality is achieved, when a node is in a brokering position, it can be used to measure Bridging Social Capital. Therefore a high value indicates that a user has a lot of connections which pass information by him. The position in the network of such a user allows one to exert influence over the flow of information. As the user can decide to withhold or pass along information, the user is in a position of power (Wasserman and Faust 1994, p. 188).

This power makes users to knowledge hubs according to Berger et al. (2014). They are required to distribute information in the network and diffuse innovative ideas to other groups of people in the network. Angeletou et al. (2011) acknowledge this and relate a high value to an *influencer* type of person. *Influencer*s are able to initiate discussions and spread information by engaging in extended conversations. The engagement characteristic is picked up by Hacker et al. (2015), who describe it as "*high levels of continuous engagement*" (p. 17). It is related to a brokering position in the network and thus constitutes for Bridging Social Capital.

Riemer et al. (2015) remark that for the brokering effect to take place and improve the individuals performance, a regular activity in the network is required. A low betweenness centrality implies a lack of Bridging Social Capital.

Eigenvector Centrality

The idea of the *eigenvector centrality* is to not only look at the number of neighbours a node has, but to also look at the importance of the node's neighbours. Therefore, a node has a high eigenvector centrality if it either has a high number of neighbours or a few, but important neighbours.

Because the eigenvector centrality is calculated for a single node, it is of ego-centric scope. The metric is proposed by Smith et al. (2009), Berger et al. (2014) and discussed in Wasserman and Faust (1994).

The calculation is based on Bonacich (1987) and explained by Newman (2010). It makes use of the adjacency matrix A.

Let $x(0)$ be the initial guess for the vector of centralities. Because it is unknown which node may be important, I assume an arbitrary value for the centralities. This value has to be equal for all elements in the vector. A common approach is to set all elements in $x(0)$ to 1.

Now the vector is multiplied with the adjacency matrix A to calculate the update value $x(1)$:

$$x(1) = A * x(0).$$

Multiplying with the adjacency matrix assigns each element in the centrality vector the sum of the centrality values of its neighbours. This can be repeated any number of times t, thus I take the matrix A to the power of t.

$$x(t) = A^t * x(0).$$

To get a final value, I let t approach infinity. For this I write $x(0)$ as a linear combination of the eigenvectors according to Newman (2010, pp. 169f):

$$x(t) = A^t \sum_i c_i v_i = \sum_i c_i K_i^t v_i = K_1^t \sum_i c_i \left[\frac{K_i}{K_1}\right]^t v_i.$$

He describes it as follows: "K_i are the eigenvalues of A, and K_j is the largest of them. Since $K_i/K_j < 1$ for all $i \neq 1$, all terms in the sum decay exponentially as t becomes large, and hence in the limit $t \to \infty$ we get $x(t) \to c_j K_i^t v_j$. In other words, the limiting vector of centralities is proportional to the leading eigenvector of the adjacency matrix. Equivalently we could say that the centrality x satisfies the formula" (Newman 2010, p. 169):

$$Ax = K_j x,$$

which is the eigenvector centrality first proposed by Bonacich (1987). Newman (2010) mentions that the centrality x_i of the node v_i is proportional to the sum of the centralities of x_i's neighbours.

This means that the eigenvector centrality can be of a high value, because of the node's neighbours importance or the node's own importance. Therefore the final formula looks like:

$$x_i = K_j^{-1} \sum A_{ij} x_j.$$

A visual explanation of the calculation was created by Dan Ryan[14].

Smith et al. (2009) claim that the eigenvector centrality is interpreted similar to the betweenness centrality. Users that form relationships with important neighbours become relevant themselves (Newman 2010, p. 169). Due to their social relationships they fill a structural hole in the network indicating Bridging Social Capital. The neighbours pass information along a user with high eigenvector centrality. Therefore, such a user sees a lot of information and can disseminate the information to more users. Berger et al. (2014) identify this as a characteristic of key users, who contribute the most to the network. A low eigenvector centrality implies that a user is not well connected with other users. Thus the user is not in a position, where one receives a lot of information, which is associated with a lack of Bridging Social Capital.

[14] http://djjr-courses.wikidot.com/soc180:eigenvector-centrality (accessed 2017-01-13).

4.2.2 Global Metrics

The following section provides a look into metrics, which are concerned with the graph network as a whole. It does not contain metrics for single nodes, but instead provides metrics over all nodes.

Number of Nodes and Edges

These two metrics count all the *nodes or edges*, respectively, in the network. Since it is not concerned with individual nodes, it is of global scope. It is proposed by Smith et al. (2009) and Viol and Hess (2016).

As a reminder the definition of the graph G is

$$G = (V, E),$$

where V is the set of nodes and E is the set of edges. The number of vertices is calculated via the cardinality of the vertices set of the graph G:

$$N_v(G) = |V|.$$

The number of edges is calculated via the cardinality of the edge set of the graph G:

$$N_e(G) = |E|.$$

It should be noted that in the calculations in section 4.2.1 the network size has also been named g. This is equal to the number of nodes N_v.

The number of nodes and edges directly represent the size of the network. While a high number of nodes equals a big userbase, a high count of edges represents high activity and interactions between users. The number of edges overlaps with the Enterprise Social Network metric *Messages created* and the total degree of the network. Viol and Hess (2016) relate a high level of activity with engaging discussions and the sharing of ideas. According to Steinfield et al. (2009) a high number of interactions and activity lead to bonding relationships and strong ties. These strong ties are the source Bonding Social Capital. Based on the Bonding Social Capital and the closure theory, it can be assumed that the network facilitates effective collaboration. It enables users to share ideas and solve problems in a collaborative manner.

However, a low number of nodes and edges indicates a low level of engagement and a lack of communication. To get reasonable analysis results, a minimum size of the network is required.

Graph Density

The *graph density* describes how closely the network is connected. The number of edges in the graph are compared to the maximum possible number of edges in the graph. Therefore it is of global scope. This metric is proposed by Smith et al. (2009), Borgatti et al. (1998) and the calculation is based on Wasserman and Faust (1994).

Let $g = |V|$ be the network size or the number of nodes in the graph. With loops being excluded this leads to the maximum number of possible undirected edges:

$$e_{max} = g * (g - 1)/2.$$

For a network with directed edges the value needs to be doubled as there can be two edges between a pair of nodes:

$$e_{max} = g * (g - 1).$$

The density is the ratio of the number of existing edges to the number of possible edges:

$$dens = \frac{|E|}{e_{max}} = \frac{|E|}{g * (g - 1)}.$$

The value ranges from 0, which is an empty graph to 1, which is a complete graph with all possible edges present.

The graph density describes how closely connected the relationships in the network are. It is directly related to Coleman's (1988) closure theory. A cohesive group has a common ground and understanding (Burt 2001) and is able to collaborate effectively. The dense network facilitates strong ties between the actors in the network and establishes Bonding Social Capital. Provided by the Social Capital and the strongly tied relationships is a sustained solidarity and trust within the group (Coleman 1990). This enables long-term project success and commitment to the organisation (Singh et al. 2011; Scott 2012). A low density shows a lack of cohesion in the network, and therefore measures a lack of Bonding Social Capital. The users are not interacting on a regular basis in the network.

Clustering Coefficient

The *clustering coefficient* describes the tendency of nodes to create tightly knit groups or cliques within a network. It is based on the number of connected triplets of nodes in the network. The metric is of global scope and is proposed by Smith et al. (2009), Hacker et al. (2015) and Viol and Hess (2016). Technically it can also be calculated for local nodes as described by Wasserman and Faust (1994). Due to the above authors only mentioning the global variant, this is ignored at this point. Its calculation is based on Wasserman and Faust (1994).

A closed triplet is three nodes i, j, k, which are connected in the form:

$$i - j, \quad j - k \quad and \quad k - i.$$

The direction of the edges are irrelevant for the triplet. A connected triplet is three nodes connected in an arbitrary form. The difference is illustrated in Figure 8[15]. On the top is the closed triplet and the bottom shows the triangle.

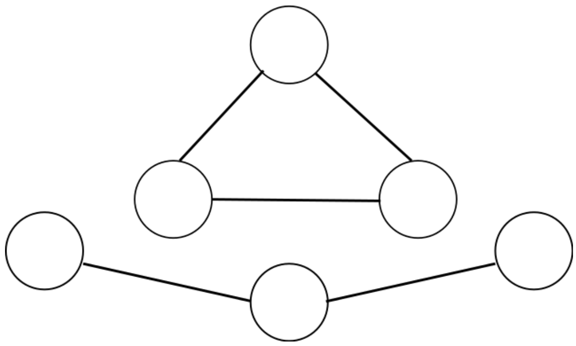


Figure 8 Clustering Coefficient

For the calculation of the clustering coefficient the number of closed triplets is divided by the number of connected triplets:

$$clustering = \frac{closed\ triplet}{connected\ triplets}$$

Smith et al. (2009) relate a high clustering coefficient to engaging users, who like to participate in discussions. However, Hacker et al. (2015) narrows this interpretation and adds that this engagement is limited to specific topical interests. Viol and Hess (2016) calls this type of user a *niche expert*, who has a tendency to interact with people from his groups, but not with people from other groups.

Thus a high value shows the existence of subgroups in the network. While they can be very cohesive within, these subgroups are able to oppose other subgroups or the general norms of the network. While Bonding Social Capital is based on strong ties and cohesive groups, it can also facilitate subcultures (Portes 1998) that stand orthogonal to the main culture (Hatch 2012). This fragmentation can lead to a decrease in productivity as a shared understanding between subgroups is lacking. A low clustering coefficient indicates a homogeneous network with low fragmentation. This is beneficial for effective cooperation in the network.

[15] https://commons.wikimedia.org/wiki/File:Triple.svg (accessed 2017-01-17).

4.3 Enterprise Social Network Metrics

The data set is coming from a structured relational database. Therefore I model the data structure in relational algebra (Codd 1970) and express the calculation schemas in structured query language (SQL).

Since the calculation schemas for the combined metrics are more than half a page long, I decided to only describe the short form in this section in pseudo-code. The full queries can be found in the appendix.

The structure of the relations are given based on the data description in chapter 3.3. Several attributes are irrelevant for the calculation schema and thus left out (see chapter 3.3 for comparison).

```
relationships(id, class, source, reciprocal, target,
              threadid, date, groupid, privacy);
nodes(id, date, state, deleteddate);
```

The following section is divided into the two subsections discussed in the metric repository structure: single metrics and combined metrics.

4.3.1 Single Metrics

In the following I look into detail at single metrics, which have been mentioned by at least two authors.

Messages Created

The first metric is the *number of messages created*. It counts all messages created by a particular user or group including all private and public messages. Since it can be calculated for individuals and groups, it is of ego-centric and global scope. It is proposed by Berger et al. (2014), Hacker et al. (2015) and Smith et al. (2009). Viol and Hess (2016) suggest a variant of this metric called "public messages" created. Because of the metrics' similarity, both of them are discussed in the following.

For the calculation of the message count m of the user with the id *uid* two steps are necessary. First, I select the data rows whose source attribute is equal to the given user id from the relationships table. Second, I run the *count* aggregate function that returns the total number of messages of the particular user:

```
m(uid) := SELECT COUNT(source) FROM relationships WHERE source = uid;
```

For the calculation of the message count *m* of a group with the id *gid*, the first step selects the data rows with the given group id. The second step stays the same and returns the total number of messages of the particular group:

```
m(gid) := SELECT COUNT(groupid) FROM relationships WHERE groupid = gid;
```

Riemer et al. (2015) strongly relate the number of posts to the level of participation of a particular user. They theorise that "*participation in ESN constitutes social capital*" (p. 9) and active users have a higher job performance than non-active users.

Berger et al. (2014) operationalise user activity via the number of posts. They state the activity is an important factor for the identification of key users. Key users are responsible for a significant number of messages. This is reasonable, because if users contribute a lot, they share their knowledge and help other users. This assumes that the posts are related to professional services of the organisation, which is the case in Enterprise Social Networks (Riemer et al. 2015).

Viol and Hess (2016) determine that a high number of created messages characterises a *power user*. Power uses are well connected (cf. degree- and closeness centrality) and highly visible. They are always online and among the most popular users of the network. Contrary, a low number of messages indicates either a *temporary user* or a *niche expert*. The former is not very active in the network and does not participate in discussion as the user usually takes a long time to reply in a conversation. Such a user receives a lot of likes and thanks messages and thus is popular and helpful towards other users.

The number of posts overlaps with the degree- and closeness centrality, which is shown by findings from Hacker et al. (2015). They correlate a high number of posts to a high engagement, but narrow focus. Smith et al. (2009) agree and find that a high number of messages characterises users, who are very active and engaging in discussions. They spawn new threads, new ideas and contribute to other threads, potentially helping other users. Angeletou et al. (2011) call such users *elitists* and *popular users*, who communicate a lot with their peers and are drivers of activity.

Because the message count is not a structural aspect of the network, it is not per se an indicator of Social Capital. However, due to the common factor and high similarity with the degree- and closeness centrality, I want to apply the internal perspective of Social Capital to the message count. Steinfield et al. (2009) state that a high number of posts and activity leads to bonding relationships and strong ties. These social relationships facilitate Bonding Social Capital, which enables a common understanding and effective collaboration. Contrary, a low number of posts indicates a lack of cohesion and therefore implies a low level of Bonding Social Capital. This interpretation is applicable to the ego-centric and the global version of this metric.

Threads, Replies and Likes Created

The metrics *threads created*, *replies created* and *likes created* are concerned with the content a user generates. I define a reply as a response to another message and a thread as a post that spawns a new discussion i.e. it is not a reply to another message. A like is a non-textual acknowledgement of a message. This metric is proposed by Smith et al. (2009) and Viol and Hess (2016) and can be calculated for individuals or the whole network. Therefore it is of ego-centric and global scope.

For the calculation of the metrics thread count *tc*, reply count *rc* and like count *lc* of the user with the id *uid* the following queries are needed:

```
tc(uid) := SELECT COUNT(source) FROM relationships
           WHERE class = "Post" AND source = uid;
rc(uid) := SELECT COUNT(source) FROM relationships
           WHERE class = "Reply" AND source = uid;
lc(uid) := SELECT COUNT(source) FROM relationships
           WHERE class = "Like" AND source = uid;
```

Viol and Hess (2016) mention that a *power user* is active and engages in discussions on a regular basis. Due to the high number of posts, the user is visible in the network and the content influences the network as a whole. A low number of posts indicates a *temporary user* or inactive user. In social networks there can be a high number of inactive users, who are registered but not active anymore.

Users that are only active in a topical niche and only respond to threads in their niche have a low post number. Their contributions benefit their niche, but not the network as a whole.

This aligns with the interpretation from Smith et al. (2009), who relate a high thread count to users that create content and are active in the community. They propose innovative ideas and spawn new discussions. A high reply count indicates a high engagement with other users and a participation in extended discussion threads. According to Smith et al. (2009) a low metric values indicate a passive user, who may consume content, but does not contribute content. There are also cases, in which the thread count is low, but the post count is high. These users do not create new threads, but take part in discussions spawned by others, where they share their ideas.

The sum of the metrics (threads, posts and likes) equals the out-degree (cf. degree centrality) of a user in the network. A high value is associated with Bonding Social Capital based on the closure theory (Coleman 1990). Strong ties between the users lead to a cohesive group, which makes effective use of resources (Burt 2001). A low value for these metrics implies a lack of Bonding Social Capital. This interpretation overlaps with *public messages created* as that metric is the sum of threads and replies created. Therefore, the interpretation is the equal to the *public messages created* metric, where the effects are described in more detail.

Registered Days

The metric *registered days* counts for how long a particular user or group is active in the network. Since it can be calculated for individuals, it is of ego-centric scope. It is proposed by Viégas (2004), Smith et al. (2009) and Viol and Hess (2016).

For the calculation of the metric *rd* of the user with the id *uid* the following query is needed:

```
rd(uid) := SELECT NOW() - date FROM nodes WHERE id = uid;
```

On the one hand a recently registered user is called *occasional user* according to Holtzblatt et al. (2013), who has not yet contributed to the network. On the other hand old users tend to have a reputation in a network and are well known. They can exert influence over other members based on their reputation. It should be noted that the user activity over time should also be considered as an old user may have gone inactive over time.

Smith et al. (2009) relate users with a short registration time to *question askers* and a low connectedness within the network. They occasionally post content, but do not engage in extended discussions. Since such users are newly registered, they do not have a lot of relationships. They still have to connect and build relationships in the network. Therefore it indicates no or a low amount of Social Capital in the network. However, early adopters of the network have a large number of social contacts and a wide array of interactions according to Scott (2012, pp. 44f). Thus they form social relationships and strong ties, which is a sign of Bonding Social Capital. As there is no information about the structure of the relationships and the user's position in the network, a statement about Bridging Social Capital is not feasible.

For an analysis of the network and its users a minimum activity is required (Riemer et al. 2015). A more in-depth look is available with the metric *user activity over time*.

Group Activity

The *group activity* metric calculates how many of the total messages were posted in groups and how many public and private groups were contributed to. It is concerned with the focus of a particular individual and its community contributions. The metric shows where the user posts his content and if the user is more active in public or private conversations. It is suggested by Hacker et al. (2015) and Viol and Hess (2016) and is of ego-centric scope.

For the calculation of the metrics public message count m_{pub} and private message count m_{priv} the following queries are needed[16]:

```
m_pub(uid) := SELECT COUNT (source) FROM relationships
              WHERE source = uid AND privacy = "Public";
m_priv(uid) := SELECT COUNT (source) FROM relationships
              WHERE source = uid AND privacy = "Private";
```

Users contributing highly to groups are bound to a subset of the Enterprise Social Network. On the one hand they contribute to the specific group of the network, but on the other hand lack activity in the rest of the network. Such users only focus on their topics of interest according to Hacker et al. (2015).

Viol and Hess (2016) describe this attribute as *focus*. Users with focus make valuable contributions to the network, but only a part of the network can benefit from the content. They are engaging in discussions with other peers in their groups, but lack relationships and interactions with other people outside of their group.

Due to the cohesion of the group and shared norms and values, group activity can be associated with high Bonding Social Capital (Coleman 1990) within the groups. It can indicate effective collaboration and knowledge sharing in groups between the actors (Riemer 2005) based on shared norms and common grounds (Nahapiet and Ghoshal 1998). A negative effect could be the establishment of subcultures, that do not communicate with the rest of the network. These subcultures can possibly stand orthogonal towards the network and hinder effective collaboration (Hatch 2012).

If you take the sum of these two metrics, it will be the total amount of messages the user has created. The number would be equal to the metric *messages created*.

4.3.2 Combined Metrics

The characteristic of the combined metric is that they are calculated from the combination of several singular metrics. A typical example are calculated averages or ratios of the other metrics.

[16] Remark: in some data sets, groups are set to private or public. If a group is set public, the metric's value for private messages will always be zero, and the other way round.

User Activity over Time

With the metric temporal concentration of creating messages, Hacker et al. (2015) and Viol and Hess (2016) propose the idea to look at when messages are created and how the *user activity* changes *over time*. Angeletou et al. (2011) describe a similar metric called the *churn rate*, which is the loss of active users over time. I adopt the former definition as it is a good indicator for a user's activity. If the activity reaches 0, it can be concluded that this user is lost, which fits the latter metric. The activity can be measured for single users and therefore is of ego-centric scope. An average can be calculated for the network and be interpreted in a global scope.

For the calculation of the user activity over time *ua*, the timestamps of the messages are aggregated on per-month basis. The term “YYYYMM” means the year and month of the given date. The result is a set of key-value pairs, where the key is the month and the value is the count of posts for that month.

```
1. select all messages from a given user
2. group the messages by date in the form of "YYYYMM"
3. ua := count the number of messages per month
```

Viol and Hess (2016) state that a high and continuous level of engagement characterises a *power user*. They contribute actively and are well connected within the network. Due to their high activity, they have a short response time and react quickly to questions and new ideas. As they are well connected, they have a high visibility and can influence the opinions of the network. Holtzblatt et al. (2013) call this type of users *active contributors*.

The users are well connected within the network with a diverse set of other users. It can be assumed that such a user has strong relationships with a subset of these other users. Since it is difficult to maintain strong relationships, it can also be assumed that such a user has acquired weak relationships. In this sense the user is associated with the creation of Bridging and Bonding Social Capital for the network. The user's weak ties provide him with new information on a regular basis while the user can actively engage and collaborate with his strongly tied contacts.

A low activity can indicate a *niche expert*, who is only active in his group or an *information seeker*. The latter type of user tends to ask questions and passively consume information, but the user is lacking interactions with other users. Therefore the user is not well connected in the network. However, the former type can be well connected to other people of his topical interest. Such a user possesses strong relationships with his peers and through that is thought to establish Bonding Social Capital.

Berger et al. (2014) claim that a high user activity is a strong indicator of a healthy community. This is confirmed by Angeletou et al. (2011), who call it *community popularity*. It characterises a network, in which users engage with high intensity and motivate others users to contribute. If the activity in the network declines, it poses a serious threat to the health of the network as a whole. Therefore Angeletou et al. (2011) recommend community managers to act in case the activity drops. Hacker et al. (2015) note that a high temporal concentration of activity indicates low engagement. Instead it is preferable to have continuous activity over time, leading to consistent engagement with the Enterprise Social Network. Hacker et al. (2015) state that this indicates a brokering position in a network according to the structural holes theory. Thus, it measures Bridging Social Capital.

Due to the measurement of both Bonding and Bridging Social Capital, this metric is relevant for finding the optimum performance of a group (Burt 2001). It should be noted that *power users* are rarely found in a social network and thus other types of users prevail (Viol and Hess 2016).

Average Time until First Reply

The *average time until first reply* is an indicator for the response time of a user. This metric can be calculated for individual users and is of ego-centric scope. It is suggested by Hacker et al. (2015) and Viol and Hess (2016).

The average time until first reply *ar* is calculated for a users, who creates the first reply in a particular thread. The median is used instead of the mean, as some values for the response time can be big outliers.

```
1. select all threads in a group or network
2. retrieve the first and second post of the thread
3. omit threads, where second post is not from given user
4. ar := calculate median time difference between first and second post
   of thread
```

Hacker et al. (2015) analyse the response time of messages, which can be described as a delay in sending messages. A high value indicates a user, who is receiving a lot of information, especially from asking questions. Therefore such a user is successfully looking for information. This is also related to the number of likes received on such topics.

Continuing this interpretation, a user who responds quickly does so by writing short replies. This indicates a user who is responding with a simple acknowledgement or thanks according to Viol and Hess (2016). Such a characterisation is typical of a passive user, who is not contributing to the network. The passivity of a user is visible in high in-degree and a low out-degree values. Because Social Capital is established by forming social relationships through reciprocal interactions, this suggests a lack of Social Capital in the network.

Contrary, engaging and active users require time to think of an appropriate response and due to their activity all over the network, cannot reply that quickly. Their participation in discussions lets them interact with other users and form social relationships. The social ties facilitate Bonding Social Capital, which enables effective collaboration.

Average Replies per Thread

The metric *average replies per thread* is an indicator of an active discussion. It is proposed by Angeletou et al. (2011), Hacker et al. (2015) and Viol and Hess (2016). It is of global scope as it is calculated for the entire network.

The average replies per thread *at* metric is calculated as an average for the network. The median is used instead of the mean, as some threads can have an unusual high count of responses.

```
1. select all posts grouped by threads
2. at := calculate median over count of posts
```

A high number of replies per thread characterises an engaging network with knowledge and ideas being discussed (Viol and Hess 2016). Such a network thrives on active users, who initiate conversations and attract answers, leading to long discussion threads.

Hacker et al. (2015) argue that a high average value of replies per thread is an indicator for a lively discussion. Therefore in a network with a high *average replies per thread*, discussions take place. It indicates an engaging base of users, who interact and exchange ideas with each other.

Angeletou et al. (2011) describe a healthy network, in which there is high participation in topic discussions, that generate lots of replies. Discussions that are not driven forward, tend to die down, resulting in a large portion of unanswered threads. This can lead to a downwards spiral, in which the activity in the community decreases continuously.

A lively exchange of ideas and knowledge is a typical attribute of the closure theory. It is grounded in the creation of strong ties between the discussion partners facilitating collaboration and group work. Therefore, a high *average replies per thread* indicates Bonding Social Capital. A low *average replies per thread* indicates the lack of engaging discussions and community interactions and thus suggests a low level of Bonding Social Capital.

Threads Creation Ratio

The *thread creation ratio* is a pair of two underlying metrics. The (1) ratio between the number of threads and the total posts and (2) the ratio between initiated threads and total threads in the network. Smith et al. (2009) call these two metrics *Verbosity* and *Initation*, while Angeletou et al. (2011) are writing about *Thread Initation Ratios*. These metrics were later picked up by Hacker et al. (2015) and Viol and Hess (2016). They are of ego-centric scope as they can be calculated for individuals, although the calculation of an average over the whole network is feasible.

The calculation of the single thread creation ratio *st* and the total thread creation ratio *tt* is straightforward and can be accomplished in one step each:

```
(1): st := select count of threads / count of posts
(2): tt := select count of initiated threads / count of all threads
```

Viol and Hess (2016) and Hacker et al. (2015) conclude that a high number of threads compared to the number of posts, is a sign of information sharing. A user with many threads is informing other users about events or other news. However, Hacker mentions that their analysis result does not support this notion for threads, which do not receive any replies. These may be unanswered questions or uninteresting posts.

Another notion presented by Viol and Hess (2016) is that a high value indicates users who share knowledge and ideas with others, spawning new discussions threads. These discussion threads contribute content and ideas to the network. This fits with the first interpretation that replies are needed in the threads.

Hansen et al. (2010) describe such users as *discussion starters* and Rowe et al. (2013) as *expert participants*, while Angeletou et al. (2011) speak of *popular initators*. Due to their creation of threads, they are usually well known in the network and have high visibility. Users with a low ratio only post occasionally and are unlikely to start their own topics. Smith et al. (2009) claim that more threads are better for the network as it indicates the generation of new ideas and discussions.

Social relationships are only formed, when other users respond to a thread. Therefore this metric alone is not sufficient to make any claims about Social Capital. However, if a particular thread gathers attention, this indicates a high level of engagement in discussions and the exchange of new ideas. This facilitates Bonding Social Capital as people exchange their thoughts and ideas to form a shared understanding.

Reply Creation Ratio

The *reply creation ration* consists of two underlying metrics: the (1) first reply created ratio and the (2) last reply created ratio. They consider whether a user has made the either the first or the last reply in a comment-chain. These metrics indicate the tendency of the users to participate in discussions and are presented in one section. Hacker et al. (2015) and Viol and Hess (2016) propose the metrics, which can be calculated on an ego-centric scope.

The calculation schema for the first reply ratio fr and the last reply ratio lr are similar. The mean is used instead of the median, as most users have a value of zero or one and no outliers are expected.

```
(1): 1. select all posts from a user
     2. foreach post:
          2a. retrieve the thread
          2b. check if it is the first reply
     3. divide the count of first replies by count of all replies:
        fr := first replies / all replies
(2): 1. select all posts from a user
     2. foreach post:
          2a. retrieve the thread
          2b. check if it is the last reply
     3. divide the count of last replies by count of all replies:
        lr := last replies / all replies
```

According to Hacker et al. (2015) a high ratio of first or last replies indicates a person, who is not engaging in discussions for an extended period of time. Instead a reply in the middle of the thread indicates an active discussion. As a *niche expert* can answer a question with only one post, this metric characterises a *niche expert* as stated by Viol and Hess (2016). Therefore low value for the first reply and last reply ratios indicate engaging users and thus is a measure of Bonding Social Capital for the individual. Bonding Social Capital establishes a common ground for the userbase. The experts can freely share their knowledge and expertise with other users. Shared knowledge and regular communication are the basis for effective collaboration in an organisation (Riemer et al. 2015).

Thread Reciprocity Ratio

The *thread reciprocity ratio* checks if threads received any replies and thus attention. Angeletou et al. (2011) propose it as the *Bi-Directional Threads Ratio*, which is later picked up by Hacker et al. (2015) As it can be calculated for individuals or the entire network it is of ego-centric and global scope.

The calculation of the thread reciprocity ratio tr is straightforward:

```
1. tr := select threads with replies / all threads
```

A high number of threads with replies indicates reciprocal interactions and engaging discussions (Hacker et al. 2015). However, Hacker adds that this does not apply to users who post announcements or events primarily, because such posts do not generate replies.

Instead it applies to users who generate discussions on innovative ideas or problem-solving as mentioned by Viol and Hess (2016). Angeletou et al. (2011) share this view as users with a high ratio contribute positively to the network. They tend to like and support their community, but focus on their own group of people, where they achieve extraordinary reciprocity. The reciprocity and feeling of belongingness is related to the establishment of Bonding Social Capital. Strong ties and cohesive groups build norms and trust, that allows users to freely interact with each other, resulting in a high reciprocity ratio.

This type of interaction is effective for collaboration and knowledge work. However, networks with a low thread reciprocity ratio are lacking the common ground and shared understanding. This shows low Bonding Social Capital and such networks could benefit from more bi-directional interactions.

Passivity

The metric *passivity* measures the active contributions of a user compared to his passive content consumption. Hacker et al. (2015) and Viol and Hess (2016) calculated this metric based on the text content i.e. whether the text contained the word “thanks”. Since there are anonymised datasets without any text content, I propose to use the number of likes compared to the number of replies. Another variant of this metric can be calculated by looking at the ratio of thread views compared to the number of replies created. Because it is calculated for individual users, the metric is of ego-centric scope.

The passivity *p* is calculated by the ratio of likes versus replies created for a given user:

```
1. p := select count of likes / count of replies
```

A passive user is driven by gaining himself benefits and described as a *consumer* by Angeletou et al. (2011). Viol and Hess (2016) describe it as a *knowledge seeker*, who tries to gain insights, but does not interact with other users. This characteristic is also shared with users who have a low activity overall in the network (cf. user activity over time). This particular kind of users is active for a temporary time period until they achieve their information needs. Hansen et al. (2010) calls them *questioners* and Viégas (2004) calls them *question askers and newcomers*.

A user’s lack of communication means that the user does not form social relationships and therefore does not contribute to the Social Capital of the network. Specifically, a like is not sufficient to form a reciprocal interaction. Organisations should be looking to motivate their people to be active in the network as active interactions are a major reason for establishing Social Capital. This is related to the metric *user activity over time*.

Reciprocity

The *reciprocity* metric determines the level of bi-directional interactions between users. Its value is calculated from dividing the in-degree by the out-degree as suggested by Viol and Hess (2016). Smith et al. (2009) and Angeletou et al. (2011) propose the same idea, but divide the in-degree by the count of replying authors or the total degree, respectively. I propose to use the definition from Angeletou et al. (2011) as it describes the metric accurately and its values are normalised in the range from zero to one. This metric is calculated for individuals and therefore it is of ego-centric scope.

The calculation for the reciprocity r is as follows:

```
1. select count of posts where the given user is the targt
2. select count of posts where the given user is the source or target
3. r := count of received posts / total posts
```

Alternatively the calculation of the reciprocity r can be done via the graph measures in-degree and total degree:

$$r = \frac{d_{in}(v_i)}{d_{in}(v_i) + d_{out}(v_i)}$$

Viol and Hess (2016) relate a high reciprocity to *power users*, who are active and engage with other users in the network on a regular basis. They contribute by driving discussions and spawning threads with novel ideas and knowledge (Smith et al. 2009). A high reciprocity improves community activity and interactions between users (Angeletou et al. 2011). It is important that users contribute with a high intensity and regularity over a longer period of time. Preferably, they are active in all parts of the network, so it is densely connected. The dense connectedness and high engagement is a sign of Bonding Social Capital in the network. Reciprocal interactions are an indicator for strong social ties and Bonding Social Capital. It enables effective collaboration based on trust and a shared understanding. This teamwork can improve a network's performance according to Burt (2001).

4.4 Less Common Metrics

In the literature more detailed metrics can be found. Especially in the works from Hacker et al. (2015) and Viol and Hess (2016) other metrics are analysed. These metrics are less common and tied to the specific data set, which is used in their respective papers. For example, they require a dataset with the actual text content of the messages and extended meta-data about attachments, job position and service line. Since the metrics are only proposed by these two papers, a diversity of interpretations is lacking. Due to the scope of this thesis the metrics are not discussed in detail at this point, but instead listed in the appendix.

4.5 Summary of Metrics

In total 63 metrics were collected for this thesis, of which 41 are described in detail, leaving 22 for the appendix. Due to overlaps those 41 metrics are reduced to 5 ego-centric graph metrics and 3 global graph metrics, as well as 4 common Enterprise Social Network metrics and 8 combined Enterprise Social Network metrics. This results in the metric repository containing 20 subsections of metrics. The distribution is illustrated in figure 9.

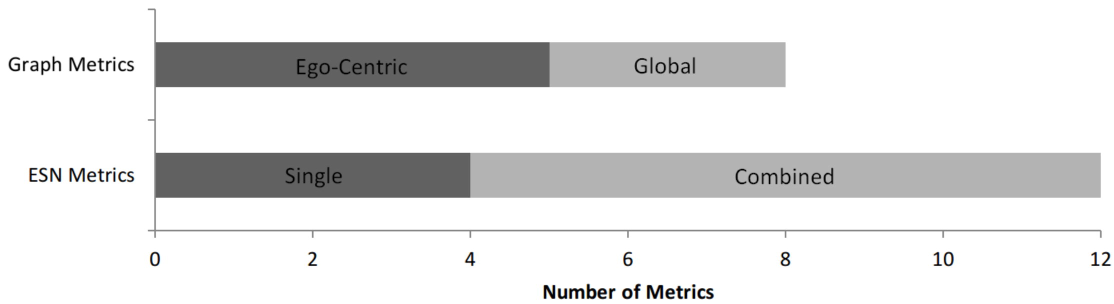


Figure 9 Metrics Number Overview

The ego-centric graph metrics contain the basic measures from the field of graph theory, that are described in the literature. Interpretations of the *degree centrality* and the *closeness centrality* overlap with the interpretation of the *eigenvector centrality*. High values for either of these measures are strong indicators for the existence of Bonding Social Capital. However, a high value for *betweenness* means that a node is filling structural holes in the network, which implies the existence of Bridging Social Capital (Wasserman and Faust 1994).

The global graph metric values measure Bonding Social Capital. *Graph density* is a typical indicator for a network's cohesion and closure (Coleman 1988). The *clustering coefficient* is suited for determining the fragmentation of a network. Thus a low *clustering coefficient* implies less fragmentation and more cohesion. The *number of nodes and edges* represents the network size. While Social Capital cannot be inferred from a high *number of nodes*, the *number of edges* equals the interaction count. Reciprocal interactions are the basis for Social Capital, so the *number of edges* quantifies Bonding Social Capital.

Both the metrics *messages created* and *posts, replies, likes created* are overlapping with the *degree centrality* metric. The former metric is the sum of the latter metric as the *posts, replies and likes created* is on a more granular level than *messages created.* The out-degree is equal to the number of messages a user created. Therefore it is natural that a high value for these metrics also implies the existence of Bonding Social Capital. In a similar vein, the *group activity* metric measures Bonding Social Capital according to Viol and Hess (2016). However, the *group activity* prominently describes possible negative effects of Bonding Social Capital as the formation of subgroups and subcultures can lead to a decline of productivity (Portes 1998; Hatch 2012). The interpretation of the metric *registered days* is based on the assumption that long-time users had a lot of time to build strong relationships and ties with other people (Holtzblatt et al. 2013; Smith et al. 2009). This results in the establishment of sustainable Bonding Social Capital.

I recommend to look at the *user activity over time* instead of the metric *registered days* as it does not include any details of the actual activity and contributions of a user. A high user activity has been associated with Social Capital in general. Due to the high level of engagement, Berger et al. (2014) and Angeletou et al. (2011) link it to Bonding Social Capital. As such an active user fills structural holes, Hacker et al. (2015) links a continuous level of activity to Bridging Social Capital.

Another metric related to time is the *average time until first reply* metric. It measures how much effort (time) a user puts into a response. If the response time is short, it is implied that the user does not put a lot of effort into his contribution. Hence, low quality content might be created that does not facilitate engaging discussions. Contrary, high quality content does facilitate engaging discussions. The effort and participation in long threads leads to the formation of social relationships and indicates Bonding Social Capital according to Viol and Hess (2016) and Hacker et al. (2015).

The metrics *average replies per thread*, *thread reciprocity ratio* and *reciprocity* are all concerned with the reciprocity of interactions in a network. *Average replies per thread* is a metric of global scope and describes the level of engagement in the network. It measures Bonding Social Capital via the amount of content and contributions per thread. Angeletou et al. (2011) argue that active discussions are required for a healthy community. The *thread reciprocity ratio* is concerned with the number of threads, that did not receive any attention and responses. Users have to initiate and take part in discussions, to form social relationships and Bonding Social Capital. *Reciprocity* is a more general metric, looking at how many posts received a response. For this metric, it does not matter if a message is a thread, post or like. In summary these three metrics are useful to to get a broad overview of the reciprocity in the network, which is linked to Bonding Social Capital.

The *reply creation ratio* metric is concerned with a user's tendency to participate in a broad range of topics across the network. It overlaps with the *clustering coefficient* as this metric is related to the fragmentation of the network. Low values for the metric indicate users who are only active in a part of the network and only engage in discussions there. They are not active in the other parts of the network. A similar metric is *threads creation ratio*. High values characterise knowledge experts as they generate topics, which receive attention and spawn discussions. This is assumed to facilitate Bonding Social Capital. Contrary, a high value for the *passivity* metric describes the opposite type of user. Such a user is only interested in the consumption of content and does not contribute to the network. Therefore the user has a lack of Social Capital.

Metric	Origin	Scope	Type of Social Capital
Degree Centrality	Graph	ego-centric	Bonding Social Capital
Closeness Centrality	Graph	ego-centric	Bonding Social Capital
Eigenvector Centrality	Graph	ego-centric	Bonding Social Capital
Graph Density	Graph	global	Bonding Social Capital
Number of Nodes and Edges	Graph	global	Bonding Social Capital
Messages Created	ESN	ego-centric	Bonding Social Capital
Posts, Replies, Likes Created	ESN	ego-centric	Bonding Social Capital
Average Time until first Reply	ESN	ego-centric	Bonding Social Capital
Reciprocity	ESN	ego-centric	Bonding Social Capital
Average Replies per Thread	ESN	global	Bonding Social Capital
Thread Reciprocity Ratio	ESN	both	Bonding Social Capital
Reply Creation Ratio	ESN	both	Bonding Social Capital
Thread Creation Ratio	ESN	both	Bonding Social Capital
Betweenness Centrality	Graph	ego-centric	Bridging Social Capital
User Activity over Time	ESN	ego-centric	Bridging and Bonding
Clustering Coefficient	Graph	global	Lack of Social Capital
Group Activity	Graph	ego-centric	Lack of Social Capital
Passivity	ESN	ego-centric	Lack of Social Capital
Registered Days	ESN	ego-centric	None

Table 2 Summarised Metrics

An overview of the metrics is shown in Table 2. They can be grouped together by their origin i.e. as graph theoretic metrics or Enterprise Social Network metrics or by their scope i.e. ego-centric or global. The interpretation is summarised in the column "Type of Social Capital". It displays what kind of Social Capital is measured by each metric.

Possible values are Bonding or Bridging Social Capital, or both, or the lack of Social Capital. In one case no reasonable effect on Social Capital could be inferred from the literature. This metric neither takes into account a user's position in the network nor the user's relationships.

While overlaps in the interpretations can be seen in Table 2 or section 4, distinct characteristics describe the different kinds of users. Although the broad interpretation with regards to Social Capital is the same, the differences can be found on the more granular level. Identified characteristics are located in either of the following areas: fragmentation and focus (Viol and Hess 2016), regular activity over time (Berger et al. 2014), sharing knowledge as discussion starters (Angeletou et al. 2011), belongingness and support (Angeletou et al. 2011), passivity and self-benefits (Hansen et al. 2010; Viégas 2004) and community health (Berger et al. 2014).

Ultimately all characteristics are attributed back to either Bonding or Bridging Social Capital, or the lack thereof. I recommend calculating all metrics and getting a broad overview of the network for making informed decisions based on the metrics' values. As the metrics are just indicators, analysing only one metric can lead to wrong conclusions. For example, a high user activity could also be caused by a spam bot instead of an engaging user.

The collection of the metric repository shows that some metrics are better researched than others. For example, the graph metrics are cited and used in more papers than the Enterprise Social Networks metrics, which were mostly gathered from Berger et al. (2014), Smith et al. (2009), Angeletou et al. (2011), Viol and Hess (2016) and Hacker et al. (2015).

Since the graph metrics directly describe social structures, it is clear how to interpret the values in the context of Social Capital. However, it is considerably more difficult to map the Enterprise Social Network metrics to Social Capital since they do not describe structural aspects. This is visible in Table 2 with the lack of effect on Bridging Social Capital as it cannot be inferred from metrics like *message created* or *user activity over time*. However, it is possible to infer cohesion and thus Bonding Social Capital from metrics like *message created* or *user activity over time*. For this I relied on the results and interpretations of the mentioned authors. The difficulty of interpretation is reflected in the shorter length of the particular sections of the metrics.

All in all, the metric repository with the calculation schemas and interpretations provides a basis for operationalising Social Capital. Based on the repository a software platform can be implemented and further research can be conducted.

5 Visualisation Prototype Design

In the following section a software prototype is designed. The goals for the prototype are to calculate the discussed metrics and provide visualisations for the calculation results. Users of the prototype should be able to view and learn about the metrics.

First, I describe the requirements and the overarching design of the prototype. After that, the backend and the frontend design are discussed. At last, reasons for the technology choice are given.

5.1 Requirements Specification

I present a use-case diagram with the desired use-cases which a user would perform. From these use-cases, I derive concrete requirements for the prototype. The requirements are split into functional requirements and non-functional requirements. Functional requirements are concerned with features of the prototype, while the non-functional requirements are about organisational or technical restrictions.

5.1.1 Use-Cases of Prototype

A typical end-user uses the software to get an overview of the network. This overview includes several actions that are illustrated in Figure 10. The overview shows high and low performing groups according to the calculated metrics. Based on the metrics' interpretation high- and low-performing groups can be identified. High-performing groups participate in engaging discussions and share their knowledge and expertise. Low-performing groups show a lack of activity and engaging discussions. The latest discussions and information exchanges are displayed and can be viewed. In general, the overview can be used to find out which groups are linked to high or low Social Capital.

Besides getting an overview of the network, a user can get detailed information about specific groups. The details include all results for the calculated metrics. Based on the metrics interpretations (cf. section 4) a user can infer conclusion about the group's Social Capital.

The prototype does not only show the metrics, but also serves as a knowledge hub. It provides background information to all the metrics, their calculation schemas and the metrics' interpretations. Users can inform themselves about each of the metrics and get a better understanding of the theory. Based on this understanding they can interpret the metrics in the context of Social Capital.

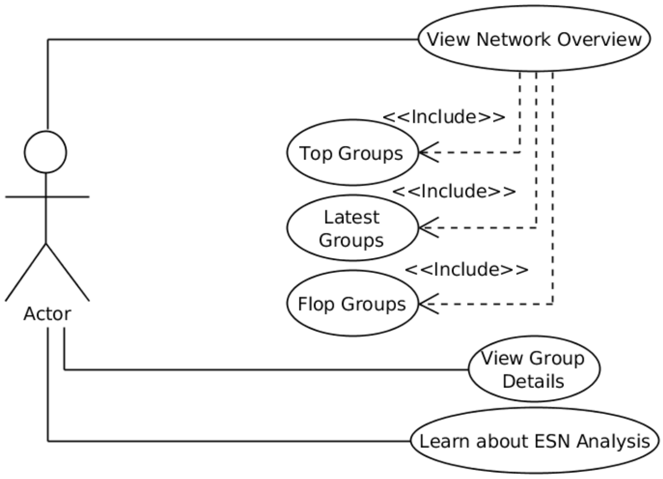


Figure 10 Use-Cases

5.1.2 Functional Requirements

To provide the metrics' calculation results in the overview and detail page, multiple software components are required. One component is the backend calculation, which is responsible for calculating the metrics from my dataset. The second component is the backend API, which takes the calculation results and exposes them to remote clients. The third component is the frontend, which aggregates and visualises the results to the end-user. Each of the components has different requirements which are outlined in the following functional requirements. CA are **CA**lculation requirements, BA are **B**ackend **A**PI requirements and VA are **V**isu**A**l requirements. An overview of the functional requirements is illustrated in Table 3.

CA 01. The prototype should be capable of calculating the graph and Enterprise Social Network metrics that were discussed in the metric repository. This excludes the less common metrics as they cannot be calculated based on my dataset.

CA 02. The graph and Enterprise Social Network metrics require different calculation approaches and the calculation of the graph metrics takes longer than the Enterprise Social Network metrics. Thus they should be calculated independently from each other.

CA 03. All calculation results should be persisted in a relational database, so they can be exposed to clients at any time without requiring a new calculation each time.

CA 04. To allow the lookup of old calculation results and the comparison of different results, a complete history of all calculation results is to be stored in the database. Old results are not overwritten, but instead all results are provided with a timestamp.

Id	Description
CA 01	Calculate all Graph and ESN metrics, except less common
CA 02	Calculate Graph and ESN metrics separately
CA 03	Provide Calculation History
CA 04	Persist all Calculation Results in Relational Database
BA 01	Separate REST API that provides the Data from Database
BA 02	Retrieve Details for All Groups
BA 03	Retrieve Results for Single Groups
BA 04	Retrieve General Information
VA 01	Overview of Groups with Appropriate Visualisation
VA 02	Utilise Appropriate Forms of Visualisation
VA 03	One Chart per Metric with Quick-Jump to Wiki

Table 3 Functional Requirements

BA 01. Remote clients should be able to retrieve the results from the database. An API should provide such access to clients.

BA 02. A user should be able to identify top-performing groups in the network based on high- and low-performing groups ranked by their metric values. The API should provide this overview data of the network.

BA 03. Besides the network overview, a user should be able to get detailed information about a single group. This includes all calculated metrics and access to the complete calculation history. The API should provide this detailed data of a group.

BA 04. Apart from the calculation results, the API should provide general information about the network and the groups such as name, id, latest posts, threads and users.

VA 01. Based on the API, a web-frontend is to be built that includes a visual representation of the network overview and group details. A user should be able to compare groups by the means of visual and tabular display of their metrics. Appropriate forms of visualisation and tables are to be chosen.

VA 02. Each metric or a combination of metrics should be visualised by one type of chart or table. This visualisation should be clearly demarcated from other metric visualisations. Each visualisation should provide a quick-link to a detailed information page of the displayed metric.

VA 03. Based on the detailed information pages of the metrics a wiki is to be built that encompasses all the details about the metrics and additional background information about the theoretical underpinnings.

5.1.3 Non-Functional Requirements

The prototype relies on the data provided by Swoop and my aim is to develop a software that is compatible to the technology utilised at Swoop. This imposes several non-functional requirements on the prototype which are illustrated in Table 4 and discussed in the following.

NF 01. As Swoop uses Highcharts.js in their Social Analysis platform, all visualised charts are to be implemented with the Highcharts.js library.

NF 02. The calculation should be runnable by a regular background job. After configuration it must be working headless, without a graphical user-interface and without any user-interaction.

NF 03. The calculation should be runnable at any time without interfering with database or the public API. This allows to recalculate all measures via a regular background job that does not impact the public API. Therefore the calculation component, the database and the API must be strictly separated.

Id	Description
NF 01	Use Highcharts.js for Charts (as Swoop uses this)
NF 02	Backend must be headless, without GUI
NF 03	Backend Calculation independent from API

Table 4 Non-Functional Requirements

5.2 Overarching Technical Design

The overarching software design provides a high-level view on the different components of the software that are required according to the requirements specification. The general structure and the connections between the components are outlined in Figure 11. For each component I describe its core design and functionality.

The three main components are: (1) The backend calculation for the metrics, (2) a REST API providing access to calculation results and (3) a web-frontend visualising the results. The database component is not self-developed, but instead a third-party application.

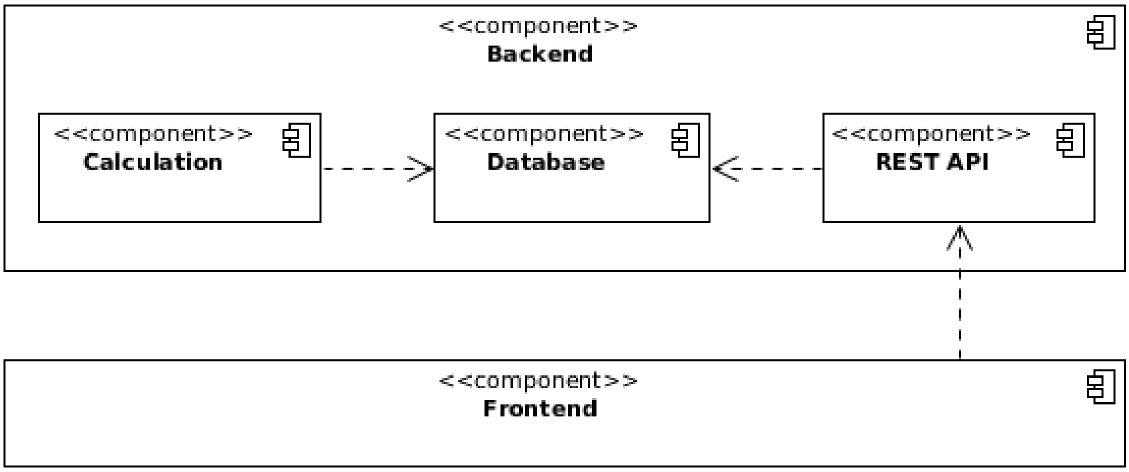


Figure 11 Software Design Overview

5.2.1 Backend Calculation

The backend calculation is responsible for preparing the database, loading and preprocessing the raw dataset, calculating the metrics for each group and persisting the results in the database.

Database Preparation and Preprocessing

Because the backend calculation processes the data and is the sole writer to the database, it has the ownership of the database. Therefore it is responsible for setting up the database, initially. This requires a `setup_db` script that creates all necessary tables, indices, PL/SQL functions and loads the raw dataset.

The following entities are included: `nodes`, `edges`, `groups`, `groups_nodes`, `threads`, `esn_analysis`, `graph_analysis` and `median(any)`. The `nodes` and `edges` hold all the raw users and raw interactions from the Swoop dataset. The `groups` relation contains the groups and the `groups_nodes` relation contains data about which user is a member of which group. The `threads` relation contains all threads that were posted in the network. The `esn_analysis` and `graph_analysis` relations hold the calculation results and the `median(any)` is a custom aggregate function required for several metrics. The complete structure of the relations is provided in the appendix with their primary keys and indices.

The raw dataset from Swoop is provided in CSV-files and is loaded into the `nodes` and `edges` tables. Since the raw dataset is anonymised the `setup_db` script provides a method called `generate_random_names()` which assigns random user names, group names and thread titles to the raw dataset. This makes the results more comprehensible for the end-user than displaying only numeric ids.

Metric Calculation and Persistence

The calculation is done in the `calc` script. It is responsible for retrieving all groups from the database, preprocessing the groups and performing the calculation. All metrics that are discussed in section 4 are calculated (**CA 01**). According to **CA 02** and because the calculation requires an external library, the `external_calc` and `internal_calc` scripts are split off of the main `calc` script.

After loading the groups, each group runs through the same process for the calculation of its values. The process sequence is performed inside a loop and illustrated in Figure 12.

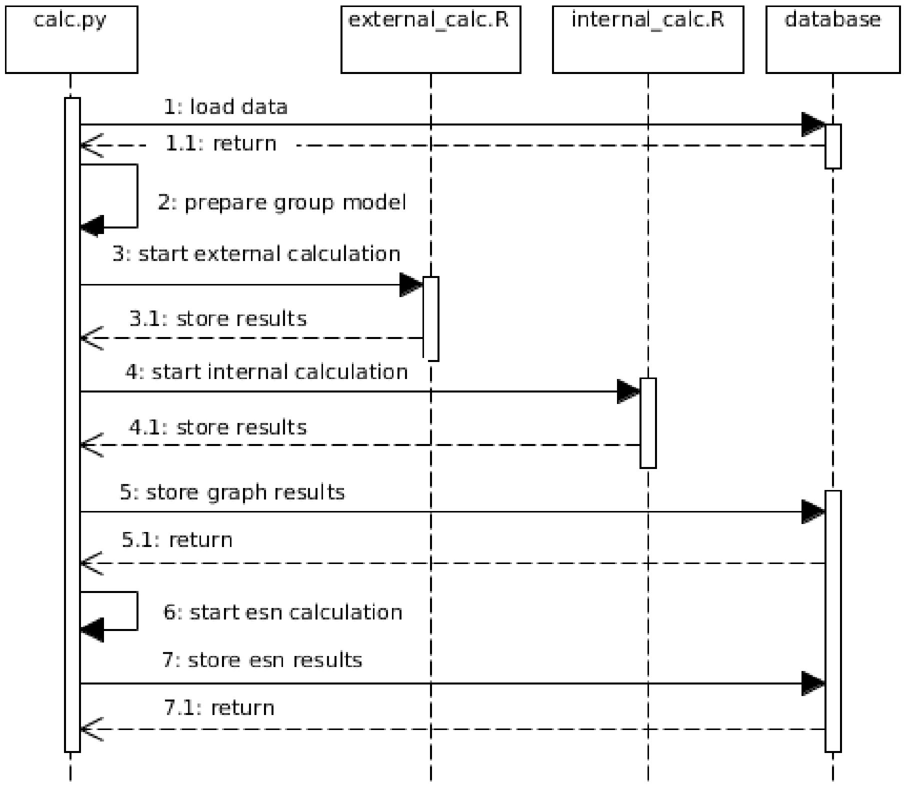


Figure 12 Group Calculation Process

At first, the group model approach is applied as proposed in section 3.5. Two versions of the group network are generated. The first version is the collective actor, which misses all internal group communication and is built from all group members and their external interactions. It is used to calculate the metrics concerned with the external perspective. The second version is built from the internal communication within a group and misses all external interactions. It is used to calculate the metrics concerned with the internal perspective.

The two versions of the network are handed off to `external_calc` and `internal_calc` scripts which calculate the graph metrics for the group and return the results. The results are persisted in the database.

After that, the Enterprise Social Network metrics are calculated with the help of SQL and PL/SQL and stored in the database (**CA 04**). Due to the complexity of the SQL statements, no ORM-wrapper is utilised. All stored calculation results are timestamped according to **CA 03**.

5.2.2 Backend REST API

To expose the calculation results to a remote client, a REST API is designed. I describe the general design of the REST API and specify the routes that were necessary to fulfil the requirements.

Structure and Components of REST API

The REST API has no direct connection to the calculation component, instead both components exchange data over the database. Several routes expose data over endpoints to remote clients. Each route consists of one endpoint, at least one database query and one mapping as illustrated in Figure 13.

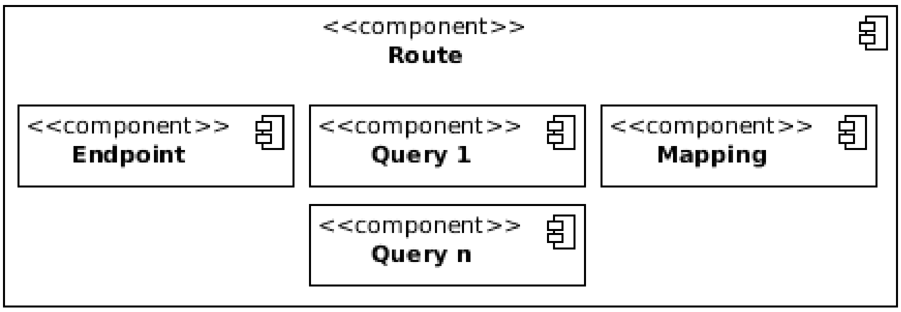


Figure 13 Route Components

The routes with their endpoints are defined in the `routes` script. An endpoint specifies the URL over which the data can be accessed and the possible parameters. A description of the route and the parameters is provided.

If the route is called from a client, the associated queries are parametrised with the client's input, so that only the requested data is returned. The queries are executed by the `queries` script and the result sets are processed by a `mapping` script. The `mapping` script transforms the SQL result sets into a unified JSON-format that is identical over the available routes. The result of the mapping is returned back to the route's endpoint which creates a

JSON response object and sends it to the client. Basically, Figure 13 can be read from left to right. It starts with the route's endpoint, executes the queries and the mapping returns the results to the endpoint.

Routes of REST API

According to the requirements specification six routes are defined and shown in Table 5.

Path	Description
/groups	Get a List of all Results for all Groups
/groups/<id>	Get the Results for one Group
/groups/deciles	Get the Deciles for all Metrics
/overview	Get Overview Data of Network
/overview/<id>	Get Overview Data of Group
/overview/heatmap	Get Activity Heatmap of Network

Table 5 REST API Routes

According to requirement **BA 01** the `/groups` route gets the latest calculation results for each group from the database. The parameters are `timeseries` and `timeseries_year`. The first parameter specifies if the timeseries data should be loaded and retrieved. If it is set to true, the second parameter specifies the year of the timeseries data to select. Since retrieving the timeseries data is not always required, the first parameter is set to false by default.

The `/groups/<id>` route fulfils requirement **BA 03** and retrieves detailed results about a group. The parameters are `id`, `calctime`, `timeseries` and `timeseries_year`. If `id` is digits-only, it specifies the group id, otherwise it specifies a search string looking for the group name. `timeseries` and `timeseries_year` are analogous to the parameters of the first route. They specify if the timeseries data should be loaded and which timeseries entry is selected. `calctime` determines which calculation results to load from the calculation history. By default the latest result is retrieved, otherwise the result which date is closest to the given `calctime` is returned. This is utilised when loading calculation results of different types. For example, a user can load the graph results from 2017-01-01 and the user can load the Enterprise Social Network results closest to the date 2017-01-01.

The `/groups/deciles` route loads the deciles[17] values for each metric from the database. The deciles are used to rank and compare groups of a single network corresponding to **BA 02**. There are no parameters.

[17] The decile is the 10-quantil used as a relative ranking for the groups.

The last requirement to fulfil is **BA 04**. The routes `/overview` and `/overview/<id>` provide general information about the network or the group, respectively. This includes the name, id, latest posts, threads and users. The `/overview/heatmap` provides an activity heatmap of the entire network showing at what times of the week there is the most activity.

5.2.3 Web Frontend

The web frontend is a website that consumes the backend REST API and displays the data in a modern and user-comprehensible web-design (**VA 01**). It is a dashboard that consists of a landing page, a group detail page and a wiki. The landing page shows the network overview and the detail page shows all calculation results for a particular group. According to **VA 03** the wiki provides background information on the theoretical underpinnings. The user interaction is restricted to the navigation of the website and the selection of different data to display. Therefore I decided to use a static website generator with templating.

The project layout with its packages is as follows:

```
/frontend
|
+-- layouts
+-- includes
+-- dashboard
+-- wiki
+-- js
+-- style
+-- node_modules
+-- dist
```

The folder `layouts` contains the main layout file that is used for all pages of the web frontend. The main layout includes the necessary stylesheets and javascript files as well as the navigation bar and the menu. The stylesheets can be found in the `style` folder and the javascript files in the `js` folder. The navigation bar and menu are designed as partials and reside in the `includes` folder. As part of the build-system the dependencies of the project are placed in the `node_modules` folder and at compile time, all resources are concatenated, minified and copied to the `dist` folder. The `dist` folder can be deployed on a web server to run the website.

The static website content and the wiki are written in markdown and converted to HTML. As interactive features are executed client-side, the dynamic functionality of the website must be implemented in javascript. When the static part of the website is loaded, the javascript requests the data from the REST API as illustrated in Figure 14.

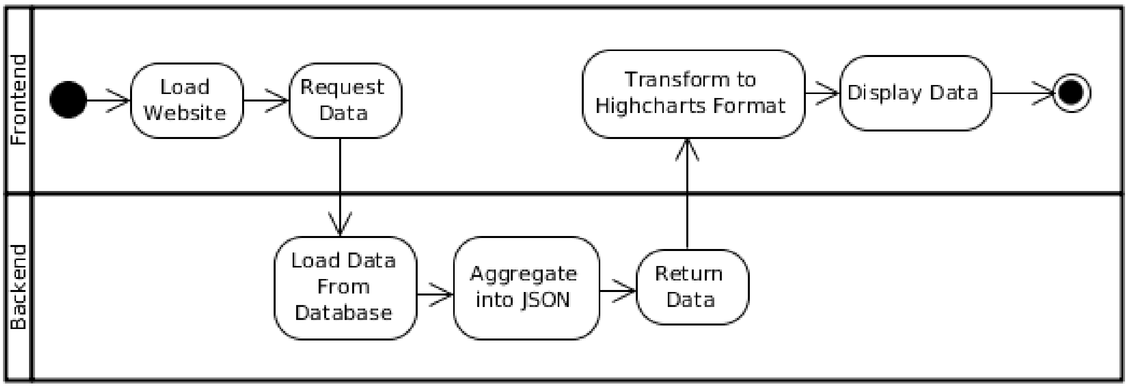


Figure 14 Web Frontend API Requests

An asynchronous request is sent to the appropriate route of the API server. This is either the `/groups` route for the overview or the `/groups/<id>` for the detail page. The other routes are used for additional information accordingly. The backend API then processes the request as explained in section 5.2.2.

The returned data is in JSON format and is converted to a highcharts-compatible format. Based on anchors in the static html (**VA 02**), the charts and tables are displayed in separate boxes. The boxes contain a header with a link to the wiki and a body which displays the chart or table.

The visual design of the web frontend is discussed in the next section 5.3.

5.3 Visualisation Framework

The visualisation framework is concerned with the visual design of the web frontend and appropriate forms of visualisation for each metric. It is based on the International Business Communication Standards (IBCS) which are practical recommendations for design in corporate communications.

The standard is used by PHILIPS, SAP, Credit Suisse Group, TU München amongst others[18] and is free to use under the Creative-Commons-License (CC BY-SA). Its goal is to enable effective management reporting and recommendations are given for the design of diagrams and tables. The guidelines are based on the three pillars: (1) content conception, (2) visual perception and (3) semantic notion.

In the following I introduce the three concepts and apply them to the available chart types of the highcharts library. From this set of charts, appropriate charts are chosen for the

[18] https://www.hichert.com/testimonials/ (accessed 2017-02-14).

visualisation of each metric. Based on the charts for the metrics the overall design of the web frontend is derived.

5.3.1 Content Conception

Content conception is concerned with the structure, order and abstraction level of content. The Minto Pyramid Principle (Minto 2003) is the approach for content conception which is taken by the IBCS standard.

As the human brain has limited process power (Miller 1956), the Minto Pyramid Principle puts an emphasis on the presentation of core statements. This is achieved by grouping and summarising arguments and ideas for the reader. By focusing on one or only few important results, distraction from the core statements is avoided. As details are only provided if needed, content can be quickly consumed by the reader.

The structure of content according to the pyramid principle is based on three rules (Minto 2003, p. 9): "(1) *Ideas at any level must summarise ideas grouped below them;* (2) *Ideas in each grouping must be same kind of idea; and* (3) *Ideas in each grouping must be logically ordered.*"

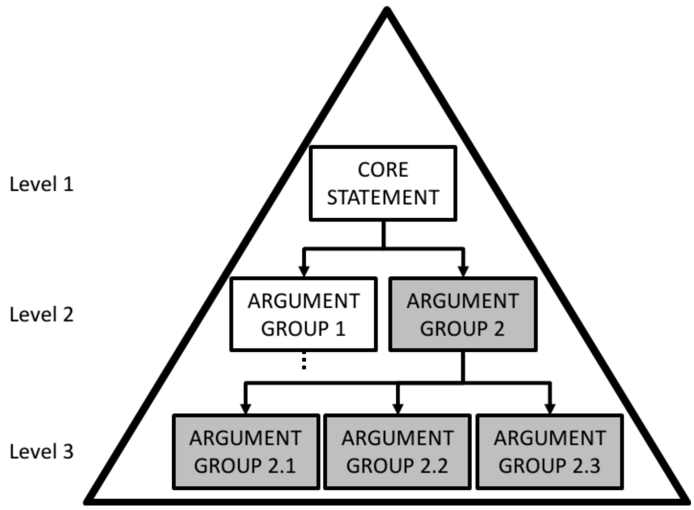


Figure 15 Minto Pyramid

In Figure 15 the structure of the pyramid is illustrated. Horizontal slices of the pyramid are called levels and vertical slices are called groups. Each argument on a particular level contains ideas and reasons for the argument on the level above it. For example, the arguments on *level two* present reasons for the core statement and the arguments on *level three* present reasons for the arguments on *level two*.

Minto (2003) insists that all arguments require a logical ordering. The ordering can be deductive, chronological, structural or comparative. Both the horizontal and the vertical order is based on this logical ordering.

The vertical order helps to capture the reader's attention (Minto 2003). A brief example for deductive ordering is given in the following:

> **Level 1**:
> "Group ABC is the best Group in the Network."
> **Level 2**:
> "Because the metrics X, Y and Z have high values."
> **Level 3**:
> "High values in X means that . . . "
> "High values in Y means that . . . "
> "High values in X means that . . . "

Level 1 is the core statement, which is explained by the argument on level 2. What the statement on level 2 means is discussed by the explanations on level 3.

The horizontal order separates the different argument groups from each other. The groups should not overlap each other and each follow different lines of thought. According to Minto (2003) two common argument groups are the answers to the questions *Why* and *How*.

The pyramid can be built by a top-down or bottom-up approach. The top-down approach starts from the core statement and builds an argumentation around the core statement. The bottom-up approach starts with detailed arguments and groups them together until a final core-statement is found. Completion of the pyramid is measured by the *mutually exclusive and collectively exhaustive* rule (Rasiel 1999). It means that there should be no duplicate arguments and no arguments missing.

In summary, the idea of the pyramid principle is to start with core statements first and go into details as needed. This is picked up in section 5.4 when the visual design is proposed.

5.3.2 Visual Perception and Semantic Notion

After discussing the structure and abstraction level of the content, it is necessary to discuss the visual representation of the content. The visual representation is divided into visual perception and semantic notion. Visual perception is concerned with the display and design of visual elements such as graphs and tables (Hichert and Gerths 2011). Semantic notion is an extension of visual perception that is concerned with unifying different types of visualisation by the means of terminology, descriptions and dimensions.

Based on Minto (2003) I want to design a dashboard that captures the most important metrics of an Enterprise Social Network. Few (2006, p. 26) provides a definition of a dashboard:

> *A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.*

He claims that charts are more efficient than texts if done correctly and thus recommends visual presentation. Users should absorb information quickly which is easier if there is only one screen. Due to the monitoring of information and frequent visits of a dashboard, the author recommends small, concise and clear display mechanisms, that clearly state their message. For this they should be customised to their specific use.

According to Few (2006) strategic dashboards show simple, high-level metrics that evaluate performance. Long-term direction is analysed and therefore they exhibit static snapshots with yearly or quarterly aggregated information. For analysts a dashboard must provide more in-depth information and comparison between metrics. Such mechanisms allow interaction with the data, choosing and comparing data by the means of different time intervals or filters. To analyse the causes for the presented information, more information should be accessible if required.

The goal is to "*squeeze a great deal of information into a small amount of space*" (Few 2006, p. 39). According to Tufte (2001) and Shneiderman (1996) it is increasingly difficult to display information as the information volume grows. Therefore user-interface designers are inventing powerful information visualisation methods (Shneiderman 1996). They use computer-aided visualisation methods to interactively engage with the information. Shneiderman (1996, p. 2) states his "Information Seeking Mantra": "*Overview first, zoom and filter, then details on demand*".

He recommends to provide an overview first and arrange the information into different boxes. Detailed information should be available with one click. Filtering, zooming and changing parameters (cf. analytical dashboard) should be possible (Shneiderman 1996).

Since the human brain seeks patterns in visualisation (Ware 2004), it is recommended to use the same style for the boxes and utilise recurring icons, colours and chart types

(Hichert and Gerths 2011). Data that is related to each other should be positioned together and similar types of data should use the same chart type (Shneiderman 1996). This allows users to make connections between chart type and data in their short-term memory which Tufte (2001) describes as "*graphics form expectations*" (p. 60). He recommends to show data variation instead of design variation, i.e. to use the same visualisation methods to display a variety of data.

The visual representation should be consistent with the numerical representation of the data, i.e. the dimensions of a chart should be proportional to the numerical quantities (Tufte 2001). In case of distorted numbers Tufte speaks of the "*Lie Factor*" which is a ratio measuring the difference between the numerical quantities and the chart dimensions.

Tufte (2001) adds that data should be displayed with clear purpose and thorough labels. Visualisation should encourage the user to think about and compare the data. It should not raise questions about the methodology of visualisation or design.

Therefore colours should be used in an appropriate amount and form (Few 2006). Excessive details or precision should be avoided, e.g. values should be rounded (Few 2006). In this context Tufte (2001) introduces the terms data-ink and non-data-ink. Data-ink are all parts of the visualisation which convey information, while non-data-ink are parts of the visualisation which are decoration or styling. Tufte suggests to minimize non-data-ink and display information with the least ink possible. Decoration and other forms of non-data-ink should be avoided.

In conclusion, a dashboard should be well organised, condensed and primarily show summaries and exceptions (Few 2006). The design is specific to the targeted audience and the information displayed (Few 2006). Complexity of single charts and graphics should be reduced, instead comparison between data points should be encouraged (Tufte 2001). The visualisation should be concise and clear in communicating its data, while maximizing the data-ink and removing redundant information (Tufte 2001).

5.4 Overarching Visual Design

Based on the visualisation framework a visual dashboard is designed for the software prototype. In the following I describe the general layout of the dashboard and reason which chart types are suited for which metrics. After the main dashboard page, the group details and the wiki page layouts are described.

5.4.1 General Layout

A mockup is illustrated in Figure 16.

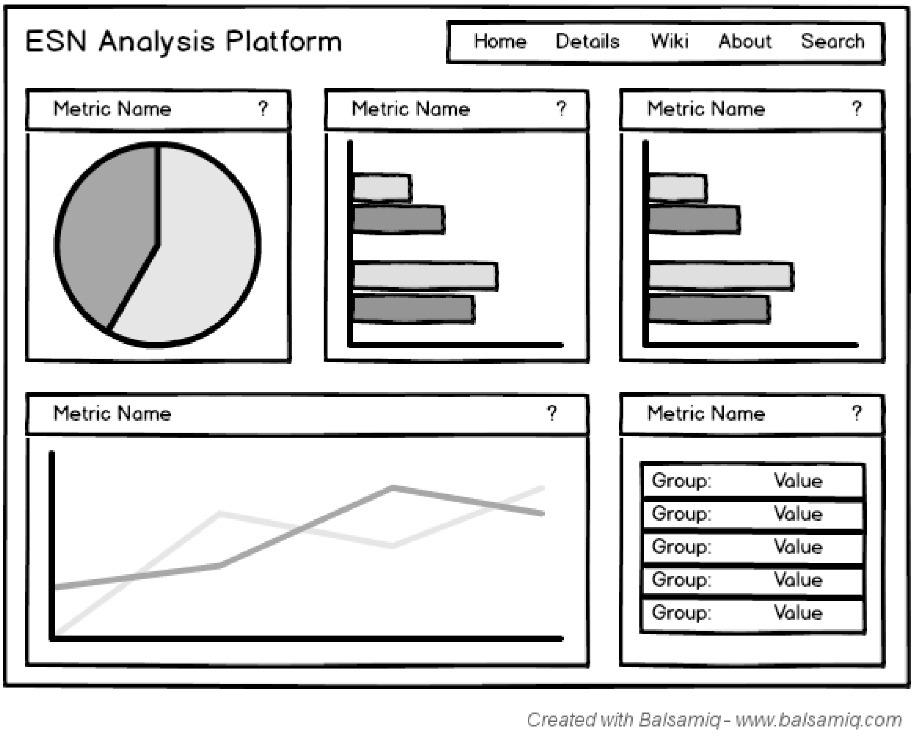


Figure 16 Dashboard Mockup

The design is split into a three column grid with a flexible number of rows. Due to my goal of displaying and educating about all metrics, I allow multiple rows and scrolling. However, the dashboard should not include too many rows, so scrolling is kept to a minimum (Few 2006).

Each of the columns contains panels (or boxes) displaying one metric or up to three metrics, if the metrics are related. The grid system allows the layout to dynamically reposition panels for different screen sizes. On a mobile screen only one column is displayed and the other columns are placed below each other. The normal size of the panels consumes one column, except for panels with big charts which can consume two columns.

The colour palette of the panels is unified with the primary colours being dark grey, light grey and blue. If information needs to be highlighted additional colours may be used. To not distract the user, all colours are from a subtle colour palette (Few 2006).

Recurring icons are the question mark which links to background information, the human icon for Enterprise Social Network metrics, the globe icon for graph metrics and the book icon for the wiki. This allows users to connect icons to a specific type of content (Tufte 2001).

Two font sizes are utilised: the normal font size which is used by default for all textual content and the smaller font size which is used for details and axis labels. The font style is unified across all pages, emphasis is accomplished by highlighting relevant texts as bold or italic. Otherwise no decorative visualisation is utilised to reduce distraction (Tufte 2001).

The panels consist of a panel header and panel body. The panel header contains the name of the metric and a quick-link to detailed information regarding this metric. This follows the pyramid principle and the recommendation of Shneiderman (1996) to display important information first and provide background information as needed.

The panel body contains a visual representation of the metric. The visual representation is limited to a specific set of layouts to avoid design variety (Tufte 2001). The layouts and charts are provided by the highcharts library and provide visual consistency. Based on Abela (2013) each type of metric is linked to an appropriate chart type. The linkages are illustrated in Table 6. In case some data does not fit in one of the given charts, it is displayed in a tabular form.

Type	Description
Line Charts	Compare Values over Time
Bar Charts	Compare Values among Items
Bubble Chart	Compare Values among Relationships between Items
Stacked Chart	Compare Values as Composition between 2 Items
Pie Chart	Compare Values as Composition of Total Share
Gauge Chart	Visualise Performance Ranking
Table	Other Values, that do not fit in Charts

Table 6 Visualisation Types

The visualisation in the panel body should provide interactive features (Shneiderman 1996). If a user clicks on a group name, he is forwarded to the group details page. In case of time data, the selection of a time interval is possible. The pie chart provides a drill-down mechanism to view detailed information about the data's composition. All charts support to hover over data points and show the exact value of a given metric. The values in the charts are rounded to three decimal digits and the values in the tables are rounded to four decimal digits.

Contrary to Tufte (2001), I display grid lines for selected charts. The bubble charts have exactly one horizontal and one vertical grid line. This divides the bubble chart into four quadrants which makes classifying the bubbles easier for a user. To let a user relate axis labels to data points, line charts and bar charts have horizontal grid lines and stacked charts have vertical grid lines. The lines are of transparent light grey, barely visible, so no distraction from the actual data occurs.

5.4.2 Mapping Metrics to Dashboard Charts

The panels in the dashboard are flexible and show different metrics as required by the network. I describe the default layout of panels that I use for the website and explain which metrics utilise which type of chart. Due to space limitations the metrics are described here and visually depicted in the appendix with screenshots.

As the metric *user activity over time* compares the activity over time, I propose to use a line chart according to Abela (2013). The data is plotted against the twelve months of the year. The user can choose which year and group to display and compare the group activity on a monthly basis.

The metrics *density*, *betweenness* and *degree* are visualised together in a bubble chart. *Density* measures Bonding Social Capital and is on the x-axis, while *betweenness* measures Bridging Social Capital and is on the y-axis. The quadrant represents Burt's (2001) optimum group performance table (cf. 2.2). The top right corner contains the groups with the highest density and betweenness, which according to Burt indicates high amount of Social Capital and group performance. The *degree* is plotted as the bubble size, so the viewer can take into account the different group sizes, when comparing the metrics.

Similar the *closeness*, *clustering* and *degree* metrics are visualised together in a bubble chart. *Clustering* is on the x-axis, *closeness* on they y-axis and *degree* is the bubble size. The top right quadrant shows the groups with the least fragmentation and highest closeness. Since the clustering measures the fragmentation, the x-axis is inverted to accomplish the correct logical order of the quadrants. The value of the *eigenvector* centrality is interpreted identical to the *closeness* centrality. Therefore the *eigenvector* centrality is not displayed to avoid redundant information. The value can be found in the table on the group details page.

To compare the *thread creation ratio* with the *average thread length* the bubble chart is used again. I propose to call this chart "interaction ratio" as it plots the number of created threads against the average thread length. The top right quadrant displays groups with a high number of threads and high average thread length, which indicates engaging users participating in the group. The *degree* is plotted as the bubble size.

The *reciprocity* compares the number of posts with replies to the total number of posts and the *thread reciprocity* compares the number of threads with replies to the total threads. The values of this ratio range from zero to one and are divided into the two parts "with replies" and "without replies". Therefore they are each depicted as stacked charts which allows users to compare the number of posts (threads) with replies and posts (threads) without replies.

The metric *group activity* compares the group's activity by the means of private and public interactions. Having two items of comparison, I propose to use a bar chart for this metric. Users can see and compare the number of private and public messages of a group.

The number of *messages created* and the respective shares of *posts, likes and replies created* of those messages is visualised using a pie chart. It shows the total composition of the messages and the different message types.

Nodes and edges are not displayed in a separate chart as they overlap with the *degree* and the *messages created* metrics. The *passivity* and *reply creation ratio* metrics are not displayed in a separate chart as they overlap with the interaction chart. Instead, these overlapping metrics can be found in a table on the group details page.

The *average time until first reply* is displayed as table in the dashboard. It shows the quickest responders in a ranking. Users can identify which groups respond to threads in a short amount of time.

5.4.3 Table and Group Details

While the goal of the landing page is to show important information and encourage comparison by visualisation and charts, the group details page makes use of tables to show all the available data. It is illustrated in Figure 17.

The group details page shows the group's ranking compared to other groups, a complete table of all metric calculation results and the calculation history. The group rankings are calculated over the deciles of a given metric from the entire network. For example, if a group has a density that falls in the top decile, then it gets a 100% ranking, if its value falls in the lowest decile it gets a 10% ranking. The ranking is visualised by gauges and supported by a subtle colour depending on the ranking-value. Red is used for bad rankings (<40%) and green for good rankings (>70%).

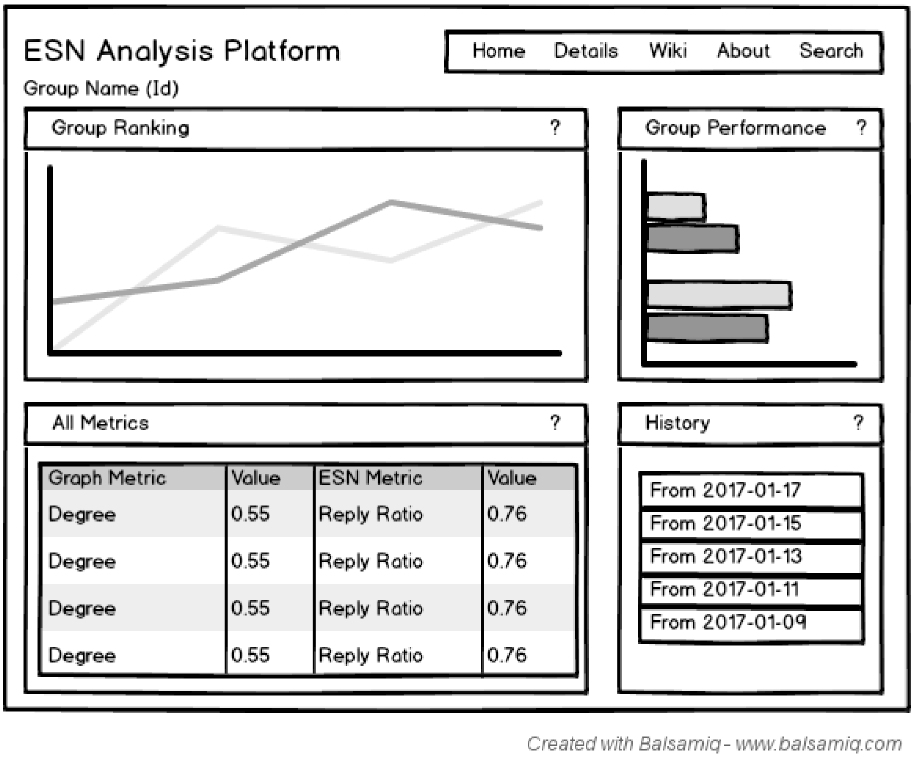


Figure 17 Groups Details Mockup

The calculation history enables the user to load old calculation results and visualise them on the group details page. By doing this a user can compare old calculation results of the group and check if the group improved over time.

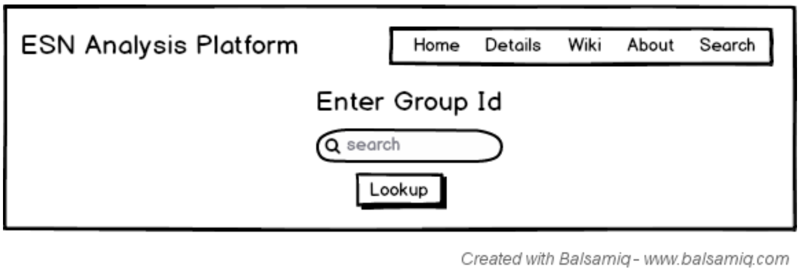


Figure 18 Groups Search Mockup

The "all metrics" table displays all metrics and their respective values for the selected calculation result. It is divided into graph metrics and Enterprise Social Network Metrics.

In case no group is selected the details page shows a search form as shown in Figure 18. A user can search a group by its numerical id or its name.

5.4.4 Knowledge Hub and Wiki

The wiki is the knowledge hub which educates the user about the theoretical background of the metrics. On the left-hand side a menu shows all the topics that are explained in the wiki.

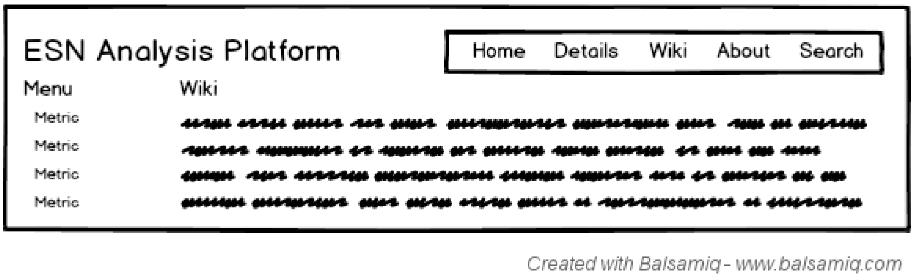


Figure 19 Wiki Mockup

They are categorised into graph metrics, Enterprise Social Network metrics and background information. The graph and Enterprise Social Network pages provide descriptions, calculation schemas and interpretations for each of the displayed metrics on the website. The background information pages provide relevant background information to Social Capital, Enterprise Social Networks and Social Network Analysis. The selected page with its content is displayed in the middle as illustrated in Figure 19.

5.5 Implementation Remarks

This section explains the technology choices and provides additional information about the deployment of the software prototype. It states where the website and the source code can be found.

5.5.1 Technology Stack

The backend calculation is programmed in Python with the database PostgreSQL. Python is commonly used in the scientific community and allows rapid prototyping. It supports scientific numbers as well as http, both of which are required for the calculation and the API, respectively. PostgreSQL is an advanced open-source database system, that is free to use and provides the necessary mechanisms for the calculation of the Enterprise Social Network metrics.

The graph metrics are calculated in R with the igraph package. The igraph package provides an implementation for the graph metrics which exactly matches my calculation schemas. Interprocess communication between R and Python is accomplished via text files.

The backend API is programmed in Python using Flask, Flask-restplus, Swagger and psycopg2. These libraries allow access to the PostgreSQL database and can be used to prototype public REST APIs. Swagger enables testing and advertising REST APIs.

Since the Flask built-in web server is a minimal server that can handle only one request at a time, the enterprise web server Nginx is used in production. It works well together with Flask using uWSGI as a middleware.

For the frontend a node environment is used which utilises assemble.io for the static website content and templating, bootstrap and font-awesome for the layout and highcharts for the chart visualisation. These technologies are chosen according to the non-functional requirement **NF 01** as they are used at Swoop. Additional software libraries include but are not limited to MathJax for formula support in the wiki, jQuery for DOM manipulation, Grunt and Npm.js for the build system and markdown for the documentation.

5.5.2 Website and Source Code

The public version of the website contains features which are beyond the design specification. These features are the latest users, groups or threads tables, the heatmap, the about page and minor quality of life changes. The latest users, groups or threads tables provide an overview of recent activity in the network. The heatmap is part of the highcharts library and provides additional insights on the user activity over time. The about page informs the user about the organisational background of the project. Minor quality of life changes in the design are not discussed explicitly.

The latest version of the website is deployed at http://www.esn-analysis.net. It runs in the Google Cloud on a virtual server with 1 vCPU and around 720MB RAM, so it is reasonably slow. The server started running on 2017-02-02 and is running until at least 2017-05-03.

If a local deployment is needed, the source code can be found on the DVD-ROM attached to the thesis distribution or in a private github-repository located at https://github.com/johuellm/masterthesis. For access to the github-repository please send an inquiry to joschka.huellmann@gmail.com or via http://www.joschka-huellmann.net. Detailed installation instructions are provided in the repository, a docker-container is planned as of 2017-02-15.

6 Discussion

This thesis set out to achieve two goals. The first is to provide an overview of metrics that can be used to analyse Social Capital. The second goal is a prototypical software that visualises the metrics on the group-level.

The operationalisation of Social Capital via Social Network Analysis as proposed by Scott and Carrington (2014) amongst others is one of the multiple ways to analyse social networks according to the Social Network Analysis framework of Stieglitz et al. (2014). In this thesis it is applied to analyse the social structures of a network (Scott 2012). Using this approach Social Capital is measured by analysing the relationships between actors and their positions in the network. Other data such as meta-data of the actors (departments, job titles) or the textual content of the interactions is not considered.

Based on different conceptions of Social Capital in the literature (Granovetter 1973; Coleman 1988), the authors Burt (2001), Adler and Kwon (2002) and Riemer (2005) point out the shared commonalities of these conceptions and propose a complementary theory of Social Capital. This theory forms the theoretical background for my thesis and analysis. Strong emphasis is put on Bonding and Bridging Social Capital and the respective perspectives (internal/external) and the theories (closure/structural holes).

I propose a metric repository that links Social Network Analysis metrics to the different Social Capital perspectives and theories which allows users to measure the effects of Social Capital. These effects include the enhancement of collaboration and cooperation, identification of information, norms and trust and the gain of individual power (Nahapiet and Ghoshal 1998; Steinfield et al. 2009; Coleman 1988; Granovetter 1973). Since Riemer et al. (2015) and Mäntymäki and Riemer (2016) claim that these effects improve group performance, management has an interest to measure these effects. Managers can use my metric repository and website prototype to analyse the effects in their network and utilise the results it in their decision-making process.

I achieved both goals and provide a comprehensive metric repository as a basis for future analysis purposes. All metrics are prototypically implemented in the website utilising the proposed group model approach to apply the metrics to groups. In the following I discuss strengths and limitations of the metric repository and further development opportunities for the prototype.

6.1 Metric Repository Discussion

For the metric repository a total of 63 metrics were collected leading to a comprehensive and diverse set of metrics. Due to the scope of this thesis and the requirements for the metric repository, not all of these metrics are discussed in detail. Specialised metrics that are not found in the majority of the literature are only discussed briefly. The specialisation of those metrics might provide unique insights that are not covered by the common metrics.

Because the metric repository is supposed to be a basis for analysis and implementation of software, the metrics are divided into categories based on their calculation schema. These two categories are graph metrics and Enterprise Social Network metrics. They allow users of the metric repository to use it as a reference for implementing software. Nevertheless, other categorisations are feasible as well. Hacker et al. (2016) categorise their metrics based on their interpretation and Riemer et al. (2015) categorise their metrics based on the metrics' scope and whether Bonding or Bridging Social Capital is measured.

The graph and the Enterprise Social Network metrics require calculation schemas which are taken from the literature. It is possible that there are different versions of calculation schemas for one metric. In this case the schema which is widely adopted in open-source implementations, e.g. igraph or tnet, is used. For the Enterprise Social Network metrics there are no schemas proposed in the literature and a lack of open-source implementations. Therefore I propose my own calculation schemas based on the description of the metrics.

All metrics are implemented in my prototype and tested against my dataset. However, further evaluation of the different calculation schemas against other datasets is recommended. This could expose optimisation potential for the schemas that can be taken into account for future versions of the metric repository. It should be noted that the interpretations depend on the chosen calculation schema and different schemas might result in novel interpretations.

The pseudo-code in section 4.3 is useful as it provides a concise and easy to understand representation of the calculation schema. However, it does not provide the fine-grained details required for a correct implementation of a given metric. Therefore, I encourage to use the SQL-statements from the appendix to implement software as they are unambiguous and provide a detailed calculation schema. The SQL-statements use advanced SQL syntax which may slightly differ depending on which SQL engine is used.

For the collection of the metrics Social Network Site and Enterprise Social Network literature was utilised. Although Richter and Riemer (2009) and Ellison et al. (2007) mention the differences between Social Network Sites and Enterprise Social Networks, both can be modelled in the same type of graph by the means of Social Network Analysis. Therefore the metrics of Social Network Sites could be adapted to Enterprise Social Networks without issues. This allows the use of metrics proposed by Smith et al. (2009) and Angeletou et al. (2011) amongst others. Besides Social Network Sites and Enterprise Social Network literature, there are research papers concerned with offline social networks by authors such as Freeman (1979) and Borgatti et al. (1998). Due to the scope of the thesis and my research approach in section 3.4, these papers are not used to identify metrics. Nevertheless, the papers are used for providing additional background information for the identified metrics.

The authors I cite provide their metric interpretation with the categorisation of low, medium and high values, that I adopted for my metric repository. What values are represented by low, medium and high depends on the particular metric and the network size. If the metric repository is utilised in an organisation for a specific network, the minimum, maximum values and quantiles should be determined to enable a comparison of the metrics.

The metrics should be seen as indicators for Social Capital and multiple metrics should be considered before deriving conclusions. The interpretations need to be discussed in the context of a particular social network (Riemer et al. 2015), so users should be careful when interpreting single metrics in their networks. As some metrics are only sparsely backed by literature, further studies can be conducted to improve the confidence in the interpretations of such metrics.

Other metrics are highly cited. They are usually “simple” to calculate and to understand and in most cases sufficient to infer conclusions about Social Capital according to Berger et al. (2014). The confidence in those conclusions is higher than in the less supported metrics that are more complex.

The metric repository is comprehensive as of 2016. The subject of Enterprise Social Network analysis is gaining traction, so more metrics might be released in the future which should be added to the repository. The repository is not a static catalogue but a document that needs regular updates to stay relevant.

6.2 Group Model Approach

In general, the global graph metrics are calculated for networks or groups and the ego-centric metrics are calculated for the individual nodes. My group model approach allows the calculation of these ego-centric graph metrics for groups. This is important for the calculation of the metrics linked to the internal and external perspectives of Social Capital.

I developed a working implementation of the group model approach in my prototype. Metrics, that were developed for use on individuals in a network, were applied to groups. To date there are no other analysis with this approach conducted, so a comparison is not feasible.

Since edges and nodes are removed in the approach, the structure of the network changes. For each group the network looks different and the side effects of the structure change are not considered. A more in-depth look into this group model approach is required to find out what side effects there might be.

For example, the graph measures are all based on the distances between nodes. By removing edges and nodes, the network artificially gets smaller and the values for these particular metrics increase. This effect is supposed to be mitigated by the normalisation in the measures which allows the comparability of particular metrics across network boundaries.

However, the number of edges and nodes that are removed varies per group. Therefore the proposed normalisation mechanisms are not sufficient and comparability of metrics across networks is not possible. A suggestion to restore this comparability would be to try and include the number of dropped edges and nodes in the calculation schema.

6.3 Implementation

The design of the prototype implementation is modular. Each of the components can be used in other projects without depending on the rest of the components. Only the dependencies must be installed for the calculation component, e.g. R and a relational database. The API can easily be extended to provide additional data, meta data and information and it can be consumed by alternative clients such as mobile applications. The frontend is responsive and works on all modern devices and browsers. Due to the usage of CSS3 legacy browsers such as Internet Explorer 11 and Firefox 48 or older are not supported.

Scott (2012, pp. 59-62f) criticises that Social Network Analysis is static and does not take into account the dynamic system that a social network is. I deal with this criticism by providing a complete calculation history. The metrics can be calculated as snapshots on a regular time basis and be compared by the analyst. Currently the frontend does not provide suitable means for comparing this data except for the user activity graph. Only the tabular display of the different calculations is feasible, but visual comparison features could be implemented in future versions of the frontend.

As my dataset is a static snapshot of a real-world database there are no changes in the data over time. Thus all calculations in the history contain the same values. Slicing the data by time intervals such as years and calculating the metrics on these slices is possible.

Another limitation of the prototype is that it is only tested against Yammer datasets as other datasets were not available to me. To use other data sources, an adapter for the database might be necessary. My data set is anonymised and lacking personal user data. To improve the visualisation I extended the data by random user names, group names and thread names. These do not reflect any real entity in the given dataset.

My dataset was lacking an explicit table with the threads in the network. Therefore I operationalised threads as posts that have gotten replies. This means that any metric which measures threads without replies cannot be calculated. The dataset provides additional classes of messages compared to the literature. While the literature only considers Posts, Replies and Likes, my dataset differentiates between Posts, Replies, Likes, Notifications and Mentions. Where only likes are considered in the metrics, the addition of notifications and mentions might be feasible.

The software prototype puts a high load on the CPU and calculation can take time. I tested the resource usage of the prototype on a laptop with 4GB RAM and a 3520M CPU @ 2.90GHz Quad Core and a virtual server with a virtual CPU and 1.7GB of RAM (Google g1-small).

The graph calculation for one group in a network with 252000 edges takes an average[19] of 1 minute on the laptop and 3 minutes on the virtual server[20]. The Enterprise Social Network measures take less than 1 second on the laptop and an average of 3 seconds on the virtual server. For a network with 1000 groups, calculating all groups would take 16.6 or 50 hours respectively.

For this test I calculated the metrics for a small network, so for bigger networks it takes more time. This raises the question of the scalability of the software. However, using a better computer with a strong CPU should be less limiting.

The bottleneck of the application is the group modelling approach as a new network graph needs to be generated for each group. Since the calculation scripts are modular and flexible, another factor mitigating the bottleneck is the partial calculation of results. The Enterprise Social Network metrics can be calculated daily and the graph metrics can be calculated weekly by the means of the calculation history.

---

[19] The average was calculated over 20 runs of the calculation.
[20] This varies depending on how much CPU time the virtual server gets.

The use-cases of the prototype are strict and minimised to limit the scope of the prototype. There are opportunities to extend the website and the analysis, but this goes beyond this thesis.

My visual design takes into account the recommendations from Shneiderman (1996) and Few (2006). However, the goal of the website is to educate and show as many metrics as possible therefore it is bigger compared to what Few (2006) suggests. Some recommendations are not implemented due to limitations in the highcharts library. While Shneiderman (1996) proposes detailed, interactive data analysis features, the highcharts library provides no advanced filtering and dynamic interactions by default.

There are several minor issues with the layout in different versions of the browsers where the CSS is not perfectly optimised. Specifically the height of the boxes in medium sized displays and old browsers is skewed and currently no user-customised positioning of dashboard elements is possible.

The metrics were mapped to the different chart types based on the characterisations of charts by Abela (2013). While the characterisations provide a guideline, alternative mappings may be feasible as well depending on the context and the purpose of the visualisation.

The software uses state of the art technologies in a modular concept with loose coupling. This makes it very flexible and usable in any constellation or software system. The disadvantage is that it requires more installation steps as compared to a monolithic software design. Therefore a docker container would make it easier for users to deploy the application[21].

6.4 Implications for Research and Practice

The metric repository provides a comprehensive starting point for research and practice alike including 63 metrics with standardised calculation schemas and metric interpretations. Due to the categorisation and concise structure of the metric repository, it can be used as reference for conducting further research or implementing software.

In research the standardisation and normalisation of the metrics' calculation schemas would allow for comparison across different papers. The group model approach allows researchers to utilise ego-centric metrics to analyse collective actors such as groups or entire networks. Instead of developing their own analysis scripts from scratch, they can use my prototype to develop software and analyse networks.

[21] A docker container is planned, but not yet implemented (cf. section 5.5.2).

Practitioners can use the prototype to see how a software system can be developed based on the available metrics. Since industry-standard technologies are used for the implementation of the prototype, it can be adopted quickly by other developers. As the software is modular and loosely coupled, customisation and adding new features is feasible without technical overhead.

The visualisation is based on best practices that come from practitioners. Decision-makers can use the website as an overview of an organisation's social network. Groups are the centres for collaboration and cooperation (Riemer et al. 2015; Bechmann and Lomborg 2012) so management is given the ability to identify trending groups, popular groups and engaging groups. The prototype is implemented as a dashboard which allows management to get a quick network overview on a regular basis. This enables the monitoring of top-ranking groups in the network. Changes in the group rankings can be observed immediately and actions can be taken in case a group's performance is decreasing. For example, management can provide premiums based on a group's ranking or try to motivate groups that are lacking engagement.

Currently the prototype is limited to the analysis on the group-level of an Enterprise Social Network. For managing an Enterprise Social Network additional tools besides the prototype should be used, e.g. analysis tools on the individual-level. The dashboard contains detailed information, so it is best used by managers with some experience in the analysis of Enterprise Social Networks. Future visualisations can provide a more simplified dashboard which also makes it usable by top-level managers without any experience in Enterprise Social Networks.

7 Conclusion

This thesis provides an overview of the latest research on Social Capital and Enterprise Social Networks. It brings these two topics together by operationalising Social Capital via Social Network Analysis. Focus is put on the levels and perspectives of Social Capital: the internal and external perspective and the resulting theories that emerged from the two perspectives. It takes Burt's (2001) optimum group performance theory as a starting point for analysing Enterprise Social Network groups. Enterprise Social Networks are discussed and compared to Social Network Sites from which they originated. It is explained why organisations are increasingly interested in Enterprise Social Networks and why research in this field is gaining traction. The effects of Enterprise Social Networks, e.g. improved collaboration and cooperation, are linked to Social Capital that is established in such networks.

Operationalising Social Capital via Social Network Analysis is done by modelling the interactions of actors in a social network as a graph where users are represented by nodes and interactions by edges. This allows the usage of Social Network Analysis metrics to measure Social Capital in Enterprise Social Networks.

A comprehensive metric repository is compiled that contains metrics and their interpretations in the context of Social Capital. Based on the metric repository a visual prototype is designed and implemented. It has a focus on automated metric calculation and appropriate visualisation based on the IBCS standard.

Future research can extend my metric repository with new metrics and existing metrics can be tested against different datasets. More studies on the metrics, like Viol and Hess (2016) or Riemer et al. (2015) conducted, can provide additional insights with regards to the metrics' interpretations. Less common metrics can be adopted to the group model approach and be implemented for the prototype. This enhances the analysis results and the comparison of groups in Enterprise Social Networks.

The group model approach has limitations that can be addressed in further research. Specifically, an approach for avoiding side-effects such as the loss of normalisation should be researched. My group model approach should be applied to other datasets independently and its performance should be compared to alternative approaches.

The analysis of meta-data information, e.g. department, job-lines, is proposed by Hacker et al. (2015). It is a trending field of research and has been applied to Social Network Sites by the New York Times and Google[22], while Viol and Hess (2016) apply it to Enter-

[22] https://jigsaw.google.com/projects/#perspective (accessed 2017-02-27).

prise Social Networks. Extending my analysis approach by the content-dimension might provide diverse insights and a new perspective on groups in Enterprise Social Networks.

In conclusion, this thesis accomplishes to successfully operationalise Social Capital via Social Network Analysis in Enterprise Social Networks. This is realised by providing a comprehensive metric repository and building a visual software prototype on top of it. The group model approach enables the use of any metric to analyse Social Capital on the group-level.

Based on the visualisation framework, presentations of the analysis results can be developed by the means of charts and tables. Such visualisations provide an overview of each group's performance in a particular network. Management can consider these group rankings in their decision-making process.

Overall, the metric repository, the group model approach and the prototype can be used by academia and practice alike. A reference of metrics including the interpretations and calculation schemas is provided for further research endeavours. For practice a basic software system is delivered that can be extended and used for the analysis of groups in Enterprise Social Networks.

With the increasing adoption of Enterprise Social Networks by organisations, their desire to analyse Enterprise Social Network data is growing. The research topic of Enterprise Social Network analysis is relevant right now and it is going to stay relevant in the near future.

Appendix

The appendix contains the followings parts:

(A) Metrics
(B) SQL Queries
(C) Software Design Database Layout
(D) Software Screenshots

A Metric List

The following is the complete list of metric evaluated. Due to space problems only the first 3 columns are displayed, namely metric description, type and scope. For the other columms such as literature source, interpretation and calculation, please consult the complete metrics.xlsx file, which is available from the master thesis DVD-ROM or from http://www.joschka-huellmann.net.

Name	Type	Subtype
GRAPH METRICS		
Degree or Degree Centrality	SNA	Ego-centric
In-Degree	SNA	Ego-centric
Out-Degree	SNA	Ego-centric
Closeness Centrality	SNA	Ego-centric
Betweenness Centrality	SNA	Ego-centric
Eigenvector Centrality	SNA	Ego-centric
Number of Nodes	SNA	Global
Number of Edges	SNA	Global
Graph Density	SNA	Global
Clustering Coefficient	SNA	Both
COMMON ESN		
Messages created	ESN	Both
Public Messages created	ESN	Both
Registered Days (or Days Active)	ESN	Both
Threads created	ESN	Both
Replies created	ESN	Both
Likes given	ESN	Both
Messages in Groups created	ESN	Both
Messages in Private Groups created	ESN	Both
Thanks created	ESN	Both
COMBINED ESN		
Temporal Concentration of creating Messages		
Churn Rate	Combined	Both
Average Time until first Reply	Combined	Global
Average Time until first Reply received	Combined	Both
Average Replies per Thread	Combined	Both
Average Replies per Thread, that Ego created	Combined	Both
First Replies created / Replies created	Combined	Ego-centric
Last Replies created / Replies created	Combined	Ego-centric
Threads created / Posts created	Combined	Ego-centric
Verbosity	Combined	Both
Initiation	Combined	Both
Thread Initiation Ratio	Combined	Ego-centric
Posts Replied Ratio	Combined	Ego-centric
Threads with at least 1 reply created / Threads created	Combined	Both
Bi-Directional Threads Ratio	Combined	Both
Messages in Groups created / Messages	Combined	Both
Private Groups Contributed / All Groups	Combined	Both
Thanks created / Replies created `use likes instead?`	Combined	Ego-centric
Outgoing	Combined	Ego-centric
Attractiveness	Combined	Both
Connectedness	Combined	Ego-centric
In-Degree Ratio	Combined	Ego-centric
Activity	Combined	Both
Community Health	Combined	Both
Bi-Directional Neighbors Ratio	Combined	Global
Reciprocity	Combined	Ego-centric
Standard Deviation of Posts per Thread	Combined	Both
Weighted Average Degree	Combined	Ego-centric

LESS COMMON		
Average Words per Message	ESN	Both
Replies to other Service Lines	ESN	Ego-centric
Replies from other Service Lines	ESN	Ego-centric
Person tags sent	ESN	Both
Topic tags sent	ESN	Both
Unique tags sent	ESN	Both
Messages with attachments created	ESN	Both
Number of Followers	ESN	Ego-centric
Average Time until first Reply to Question	ESN	Both
Average Time until first Reply to Question received	ESN	Both
Average Words per Question	ESN	Both
Replies to other Service Lines / Replies created	Combined	Ego-centric
Replies fromother Service Lines / Replies received	Combined	Ego-centric
Replies to other Job titles / Replies created	Combined	Ego-centric
Replies from other Job titles / Replies received	Combined	Ego-centric
Replies to other locations / Replies creates	Combined	Ego-centric
Replies from other locations / Replies received	Combined	Ego-centric
Questions created / Replies created	Combined	Ego-centric
Person tags sent / Messages created	Combined	Ego-centric
Person tags received / Replies received	Combined	Ego-centric
Messages with attachments created / Messages created	Combined	Ego-centric

B SQL Queries

The following contains the SQL Queries for the calculation of the ESN metrics.

B.1 User Activity over Time

The months with 0 posts are omitted in the result set.

```
SELECT to_char(effectivedatesql, 'YYYYMM') as yearmonth, count(*)
FROM edges
WHERE groupid = {groupid}
AND effectivedatesql BETWEEN '2014-01-04 00:00:00' AND '2016-12-31
    23:59:59'
GROUP BY yearmonth
ORDER BY yearmonth DESC;
```

B.2 Average Time until First Reply

There are quite many entries in the database where likes or replies are posted in the same second by other users, These are filtered out before the calculation, as this would lead to many zeros.

MEDIAN is used instead of MEAN because the value Average Time until First Reply can have very big outliers.

```
SELECT MEDIAN(d)
FROM (
  -- Selects the difference between first and second posts
  SELECT threadid, max(effectivedatesql) - min(effectivedatesql) as d
  FROM (
    -- gives a rank to based on the effectivedatesql order and it is
       partitioned by threadid
    -- thus we order all entries per thread
    -- rank and dense, so duplicates get the same rank and there are no
        gaps
    SELECT threadid, effectivedatesql, dense_rank() over(partition by
       threadid order by effectivedatesql asc) as rank
    FROM edges
    WHERE groupid = {groupid}
  ) as foo
  -- Only first post (thread itself) and the first reply are required,
     filter everything else
  WHERE rank <= 2
  -- group per Thread so we can calculate the time difference between
     thread and first reply, duplicates are removed
  GROUP BY threadid
  -- Ignore threads without replies
  HAVING max(effectivedatesql) > min(effectivedatesql)
) AS foobar;
```

B.3 Average Replies per Thread

Median is used as Average Replies per Thread can have very huge outliers.

```
SELECT MEDIAN(cntquery.count)
FROM (
  SELECT CAST(COUNT(*) AS numeric) AS count
  FROM edges
  WHERE groupid = {groupid}
GROUP BY edges.threadid) AS cntquery;
```

B.4 Thread Creation Ratios

```
-- (1) Posts / Threads
posts   := SELECT COUNT(*) FROM edges
           WHERE groupid = {groupid}
           AND class IN ('Post','Reply');
threads := SELECT COUNT(*)
           FROM (SELECT DISTINCT threadid FROM edges) AS temp;
ratio := threads / posts;

-- (2) Threads in Group / Total Threads in Network
threads     := SELECT COUNT(*)
               FROM (SELECT DISTINCT threadid FROM edges) AS temp;
all_threads := SELECT COUNT(*)
               FROM (SELECT DISTINCT threadid FROM edges) AS temp;
ratio := threads / all_threads;
```

B.5 Reply Creation Ratios

Users without First Replies are the majority and are filtered out (otherwise mean/median would be (close to) zero). MEAN is used instead of MEDIAN as most remaining users have a value of 1 and if median is 1 for every group, it is not very helpful.

(1) First Reply Ratio

```
SELECT AVG(ratio) FROM (
  SELECT first_reply_query.source, total_post_query.source, count,
      total, (count/total) AS ratio from (
    -- Selects the count per source
    SELECT source, count(rank)::numeric as count
    FROM (
      -- gives a rank to based on the effectivedatesql order and it is
         partitioned by threadid
      -- thus we order all entries per thread
      -- dense, so duplicates get removed and no there are no gaps
      SELECT DISTINCT source, threadid, dense_rank() over(partition by
         threadid ORDER BY effectivedatesql ASC) AS rank
      FROM edges
      WHERE groupid = {groupid}
    ) AS foo
    -- Only first replies are counted
```

```
    WHERE rank = 2
    -- group per source so we can get the count per source
    GROUP BY source
  ) AS first_reply_query
  -- join with self to add total posts of user
  LEFT JOIN (
    SELECT source, COUNT(*) AS total
    FROM edges
    WHERE groupid = {groupid}
    GROUP BY source
  ) AS total_post_query
  ON first_reply_query.source = total_post_query.source
) AS foobarbaz;
```

(2) Last Reply Ratio

```
SELECT AVG(ratio) FROM (
  -- select the ratio of last reply by all reply based on source
  SELECT first_reply_query.source, total_post_query.source, count,
     total, (count/total) AS ratio from (
    -- select the count of last reply per user
    SELECT source, count(rank)::numeric AS count
    FROM (
      -- select the last replies
      SELECT DISTINCT ON(threadid) source, threadid, rank
      FROM (
        -- calculate the ranks, see above for explanation
        SELECT DISTINCT source, threadid, dense_rank() over(partition
           by threadid order by effectivedatesql ASC) AS rank
        FROM edges
        WHERE groupid = {groupid}
      ) AS foo
      -- filter the post itself and posts in the same second
      WHERE rank > 1
      ORDER BY threadid ASC, rank DESC
    ) AS foobar
    GROUP BY source
    ) AS first_reply_query
  -- join with self and get total posts as well
  LEFT JOIN (
    SELECT source, COUNT(*) AS total
    FROM edges
    WHERE groupid = {groupid}
    GROUP BY source
  ) AS total_post_query
  ON first_reply_query.source = total_post_query.source
) AS foobarbaz;
```

B.6 Thread Reciprocity Ratio

This SQL also includes responses that were posted in the same second.

```
threads := SELECT COUNT(*)
           FROM (SELECT DISTINCT threadid FROM edges
             WHERE groupid = {groupid}) AS foo;
```

```
with_replies := SELECT COUNT(*)
                FROM (SELECT threadid, COUNT(*) as count FROM edges
                  WHERE groupid = {groupid}
                  GROUP BY threadid) AS foo
                WHERE count > 1;

ratio := with_replies / threads;
```

B.7 Passivity

```
messages := SELECT COUNT(*) FROM edges WHERE groupid = {groupid};
likes := SELECT COUNT(*) FROM edges WHERE groupid = {groupid} AND class
     = 'Like';
passivity := likes / messages;
```

B.8 Reciprocity

```
indegree := SELECT COUNT(*) FROM edges WHERE target = {userid};
degree := SELECT COUNT(*) FROM edges
          WHERE source = {userid} OR target = {userid};
reciprocity := indegree / degree;
```

C Software Design Database Layout

The following contains the database layout of the prototypical software. Included are all tables, indices and functions.

C.1 Edges Table

The edges table contains the raw interaction data from Swoop's dataset.

```
CREATE TABLE public.edges
(
  id integer NOT NULL,
  class character varying(32),
  source integer NOT NULL,
  isreciprocal character varying(32),
  relationshipid text,
  target integer,
  messageid integer,
  repliedtomessageid integer,
  threadid integer,
  message text,
  effectivedate text,
  groupid integer,
  privacy character varying(32),
  effectivedatesql timestamp without time zone,
  updatetime text,
  CONSTRAINT edges_pkey PRIMARY KEY (id)
)
```

C.2 Nodes Table

The nodes table contains the raw user data from Swoop's dataset.

```
CREATE TABLE public.nodes
(
  id integer NOT NULL,
  email text,
  department text,
  departmentid integer,
  name text,
  image text,
  effectivedate text,
  effectivedatesql timestamp without time zone,
  updatetime text,
  state character varying(32),
  deleteddate text,
  deleteddatesql timestamp without time zone,
  CONSTRAINT company_pkey PRIMARY KEY (id)
)
```

C.3 Groups Table

The groups table contains the aggregated group data which is generated from Swoop's dataset.

```
CREATE TABLE public.groups
(
  groupid integer,
  count bigint,
  name character varying(64)
)
```

C.4 Threads Table

The threads table contains the aggregated thread data which is generated from Swoop's dataset.

```
CREATE TABLE public.threads
(
  id integer,
  name character varying(64)
)
```

C.5 Groups-Nodes Table

The groups-nodes table contains which user is a member of which group. It is aggregated from Swoop's interactions.

```
CREATE TABLE public.groups_nodes
(
  node integer,
  groupid integer
)
```

C.6 ESN-Analysis Table

The esn-analysis table contains the calculation results of the Enterprise Social Network metrics.

```
CREATE TABLE public.esn_analysis
(
  groupid integer NOT NULL,
  calctime timestamp without time zone NOT NULL,
  messages_created integer,
  posts_created integer,
  replies_created integer,
  likes_created integer,
  notification_created integer,
```

```
  mention_created integer,
  average_time_first_reply interval,
  reciprocity numeric,
  average_replies_per_thread numeric,
  thread_reciprocity_ratio numeric,
  reply_creation_ratio_first numeric,
  reply_creation_ratio_last numeric,
  thread_creation_ratio numeric,
  thread_creation_ratio_total numeric,
  user_activity_over_time integer[],
  group_activity_public numeric,
  group_activity_private numeric,
  passivity numeric,
  registered_date timestamp without time zone,
  CONSTRAINT esn_analysis_pkey PRIMARY KEY (groupid, calctime)
)
```

C.7 Graph-Analysis Table

The graph-analysis table contains the calculation results of the graph metrics.

```
CREATE TABLE public.graph_analysis
(
  groupid integer NOT NULL,
  calctime timestamp without time zone NOT NULL,
  degree numeric,
  indegree numeric,
  outdegree numeric,
  closeness numeric,
  betweenness numeric,
  eigenvector numeric,
  density numeric,
  clustering numeric,
  nodes numeric,
  edges numeric,
  CONSTRAINT graph_analysis_pkey PRIMARY KEY (groupid, calctime)
)
```

C.8 Indices

Besides the automatically generated pkey indices, the following indices are generated:

```
CREATE INDEX idx_edges ON edges (source);
CREATE INDEX idx_groupid ON edges (groupid);
CREATE INDEX idx_groups ON groups (GroupID);
CREATE INDEX idx_threads ON threads (id);
```

The indices `idx_edges` and `idx_groupid` speed up the calculation time of the Enterprise Social Network metrics. The other two indices speed up the lookup of group and thread names.

C.9 Aggregate Functions

To prevent outliers skewing with the mean value of a set of data, the aggregate function `Median(any)` is added as an alternative to the default `AVG(any)` agregate function.

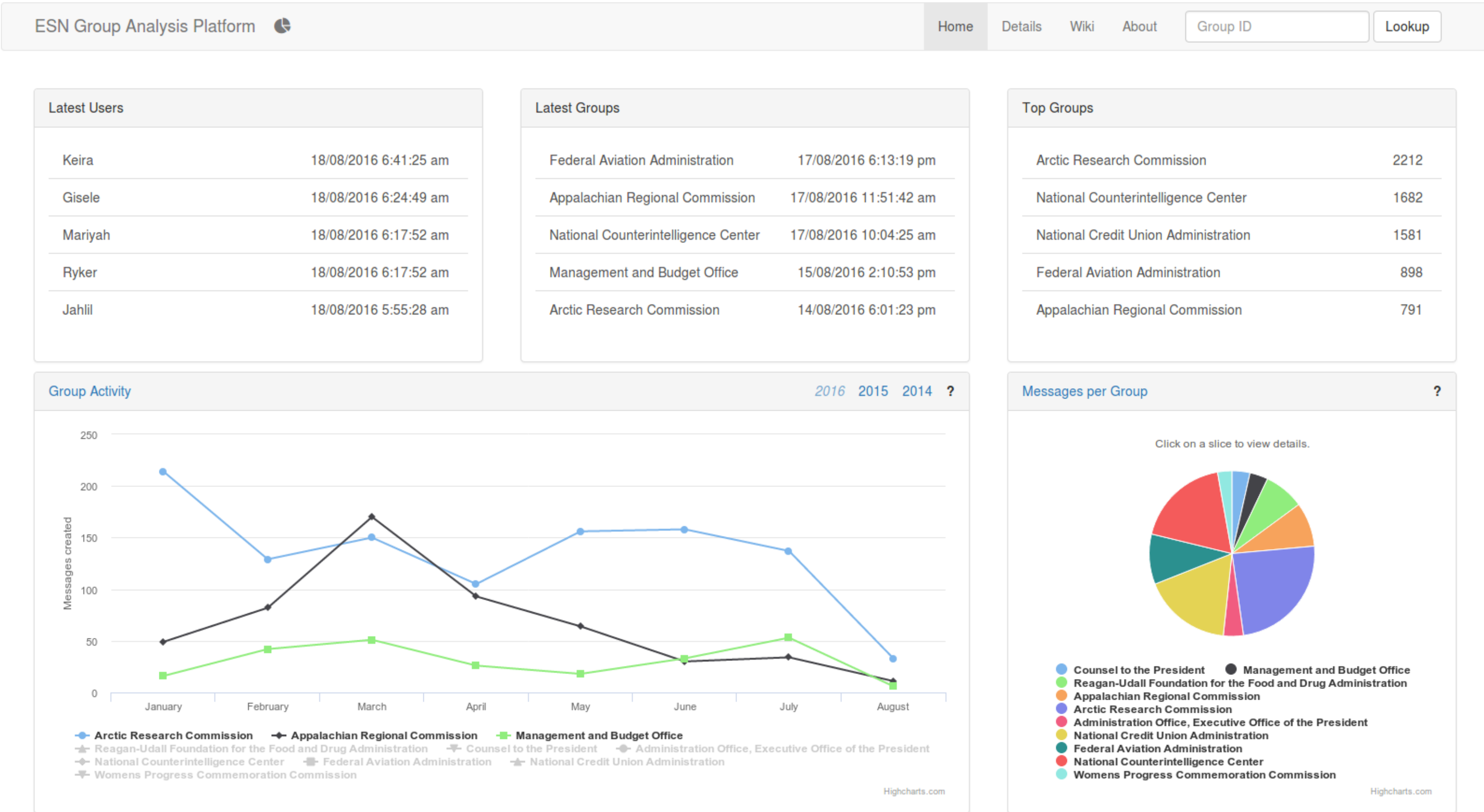
```
CREATE OR REPLACE FUNCTION _final_median(anyarray) RETURNS anyelement
   AS $$
  WITH q AS
  (
     SELECT val
     FROM unnest($1) val
     WHERE VAL IS NOT NULL
     ORDER BY 1
  ),
  cnt AS
  (
    SELECT COUNT(*) AS c FROM q
  )
  SELECT AVG(val)::anyelement
  FROM
  (
    SELECT val FROM q
    LIMIT  2 - MOD((SELECT c FROM cnt), 2)
    OFFSET GREATEST(CEIL((SELECT c FROM cnt) / 2.0) - 1,0)
  ) q2;
$$ LANGUAGE SQL IMMUTABLE;

CREATE AGGREGATE median(anyelement) (
  SFUNC=array_append,
  STYPE=anyarray,
  FINALFUNC=_final_median,
  INITCOND='{}'
);
```

D Software Screenshots

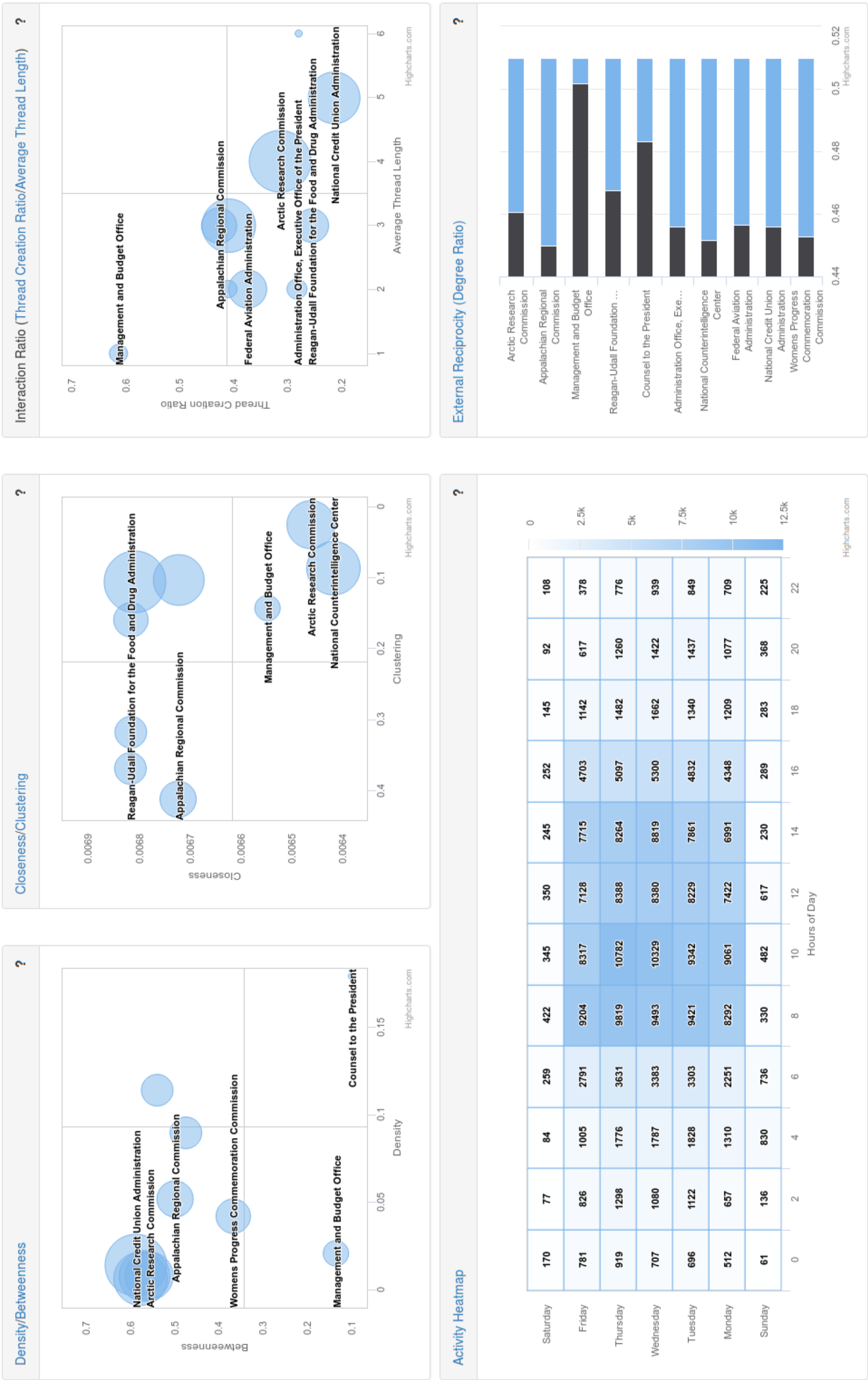
The following contains screenshots of the visual layout of the software prototype.

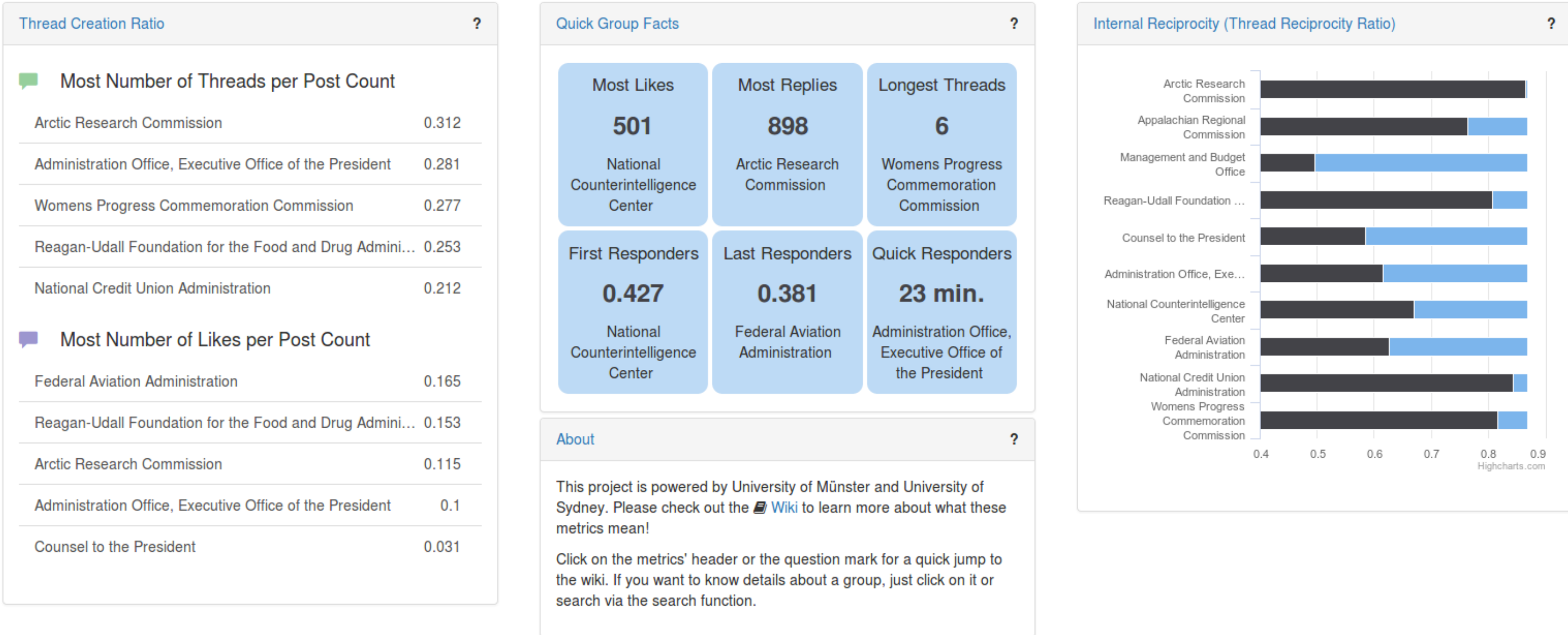
D.1 Main Dashboard

Screenshots of the dashboard.



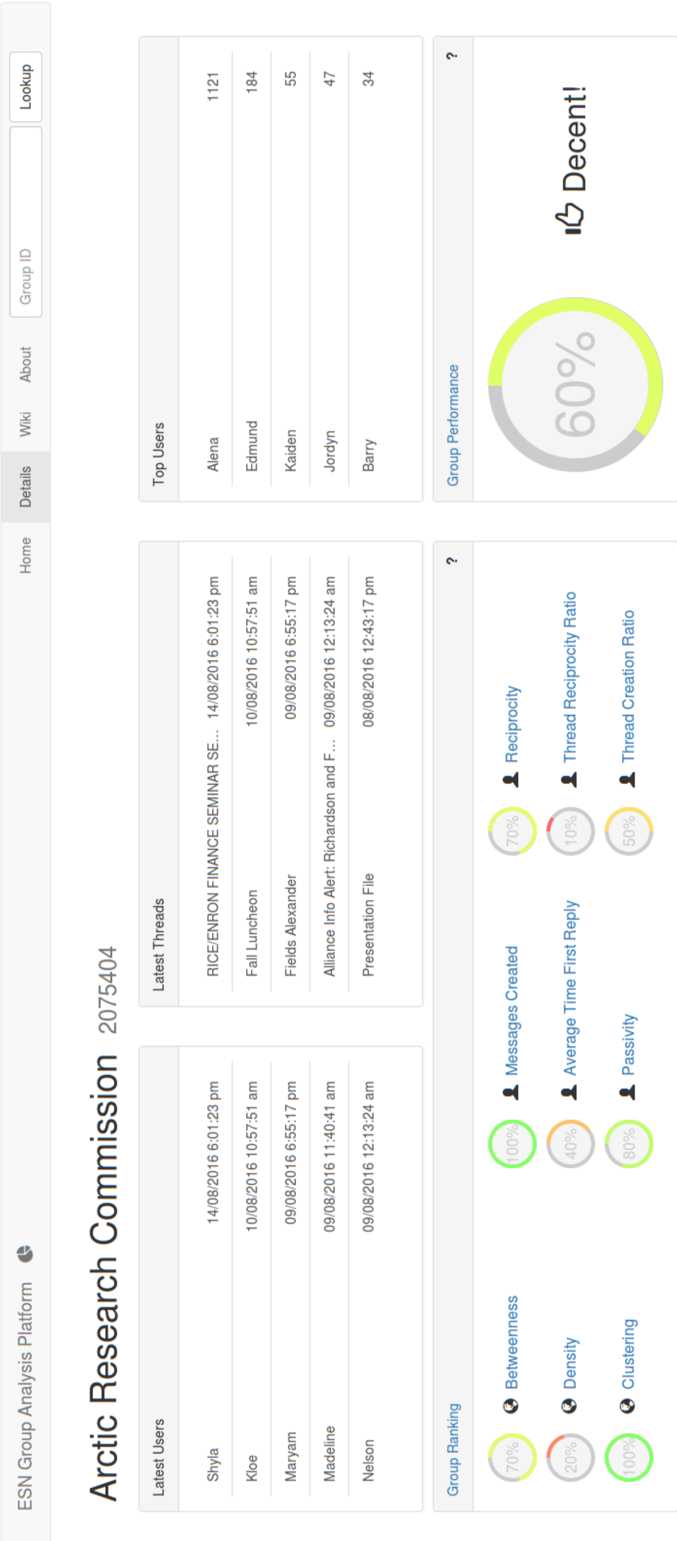
ESN Group Analysis Platform
Home
Details
Wiki
About
Group ID
Lookup
Latest Users
Keira 18/08/2016 6:41:25 am
Gisele 18/08/2016 6:24:49 am
Mariyah 18/08/2016 6:17:52 am
Ryker 18/08/2016 6:17:52 am
Jahlil 18/08/2016 5:55:28 am
Latest Groups
Federal Aviation Administration 17/08/2016 6:13:19 pm
Appalachian Regional Commission 17/08/2016 11:51:42 am
National Counterintelligence Center 17/08/2016 10:04:25 am
Management and Budget Office 15/08/2016 2:10:53 pm
Arctic Research Commission 14/08/2016 6:01:23 pm
Top Groups
Arctic Research Commission 2212
National Counterintelligence Center 1682
National Credit Union Administration 1581
Federal Aviation Administration 898
Appalachian Regional Commission 791
Group Activity
2016 2015 2014 ?
Messages created
250
200
150
100
50
0
January
February
March
April
May
June
July
August
Arctic Research Commission
Appalachian Regional Commission
Management and Budget Office
Reagan-Udall Foundation for the Food and Drug Administration
Counsel to the President
Administration Office, Executive Office of the President
National Counterintelligence Center
Federal Aviation Administration
National Credit Union Administration
Womens Progress Commemoration Commission
Highcharts.com
Messages per Group
?
Click on a slice to view details.
Counsel to the President
Management and Budget Office
Reagan-Udall Foundation for the Food and Drug Administration
Appalachian Regional Commission
Arctic Research Commission
Administration Office, Executive Office of the President
National Credit Union Administration
Federal Aviation Administration
National Counterintelligence Center
Womens Progress Commemoration Commission
Highcharts.com


Interaction Ratio (Thread Creation Ratio/Average Thread Length)
Thread Creation Ratio
Average Thread Length
Management and Budget Office
Appalachian Regional Commission
Federal Aviation Administration
Arctic Research Commission
Administration Office, Executive Office of the President
Reagan-Udall Foundation for the Food and Drug Administration
National Credit Union Administration
External Reciprocity (Degree Ratio)
Arctic Research Commission
Appalachian Regional Commission
Management and Budget Office
Reagan-Udall Foundation ...
Counsel to the President
Administration Office, Exe...
National Counterintelligence Center
Federal Aviation Administration
National Credit Union Administration
Womens Progress Commemoration Commission
Closeness/Clustering
Closeness
Clustering
Reagan-Udall Foundation for the Food and Drug Administration
Appalachian Regional Commission
Management and Budget Office
Arctic Research Commission
National Counterintelligence Center
Activity Heatmap
Saturday
Friday
Thursday
Wednesday
Tuesday
Monday
Sunday
Hours of Day
Density/Betweenness
Betweenness
Density
National Credit Union Administration
Arctic Research Commission
Appalachian Regional Commission
Womens Progress Commemoration Commission
Management and Budget Office
Counsel to the President
Highcharts.com


Thread Creation Ratio
Most Number of Threads per Post Count
Arctic Research Commission 0.312
Administration Office, Executive Office of the President 0.281
Womens Progress Commemoration Commission 0.277
Reagan-Udall Foundation for the Food and Drug Admini… 0.253
National Credit Union Administration 0.212
Most Number of Likes per Post Count
Federal Aviation Administration 0.165
Reagan-Udall Foundation for the Food and Drug Admini… 0.153
Arctic Research Commission 0.115
Administration Office, Executive Office of the President 0.1
Counsel to the President 0.031
Quick Group Facts
Most Likes 501 National Counterintelligence Center
Most Replies 898 Arctic Research Commission
Longest Threads 6 Womens Progress Commemoration Commission
First Responders 0.427 National Counterintelligence Center
Last Responders 0.381 Federal Aviation Administration
Quick Responders 23 min. Administration Office, Executive Office of the President
About
This project is powered by University of Münster and University of Sydney. Please check out the Wiki to learn more about what these metrics mean!
Click on the metrics' header or the question mark for a quick jump to the wiki. If you want to know details about a group, just click on it or search via the search function.
Internal Reciprocity (Thread Reciprocity Ratio)
Arctic Research Commission
Appalachian Regional Commission
Management and Budget Office
Reagan-Udall Foundation …
Counsel to the President
Administration Office, Exe…
National Counterintelligence Center
Federal Aviation Administration
National Credit Union Administration
Womens Progress Commemoration Commission
0.4
0.5
0.6
0.7
0.8
0.9
Highcharts.com

D.2 Group Details

Screenshot of the group details page.



Group Details

Graph Metrics
Fri, 03 Feb 2017 11:53:13 GMT

Degree	0.455
In-Degree	0.2096
Out-Degree	0.2455
Closeness	0.0065
Betweenness	0.5552
Eigenvector	1
Density	0.008
Clustering	0.025
Nodes	274
Edges	601

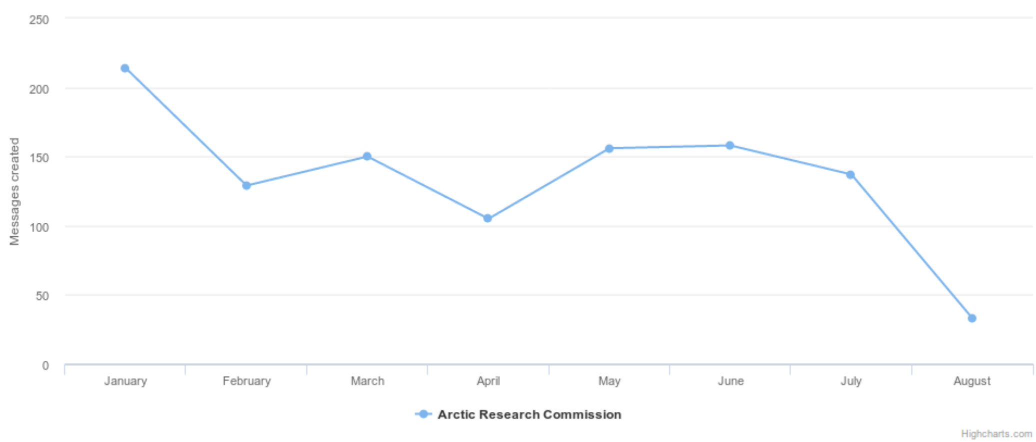
ESN Metrics
Fri, 03 Feb 2017 11:53:13 GMT

Messages Created	2212
Posts Created	407
Replies Created	898
Likes Created	255
Notifications Created	215
Mentions Created	437
Average Time First Reply	03:26:08
Reciprocity	0.4605
Average Replies per Thread	4
Thread Reciprocity Ratio	0.8649

ESN Metrics
continued

Reply Creation Ratio First	0.3544
Reply Creation Ratio Last	0.3907
Thread Creation Ratio	0.3119
Thread Creation Ratio Total	0.0069
Group Activity Public	2212
Group Activity Private	0
Passivity	0.1153
Registered Date	Wed, 19 Jun 2013 21:11:27 GMT
User Activity over Time	see graph

Group Activity 2016 2015 2014

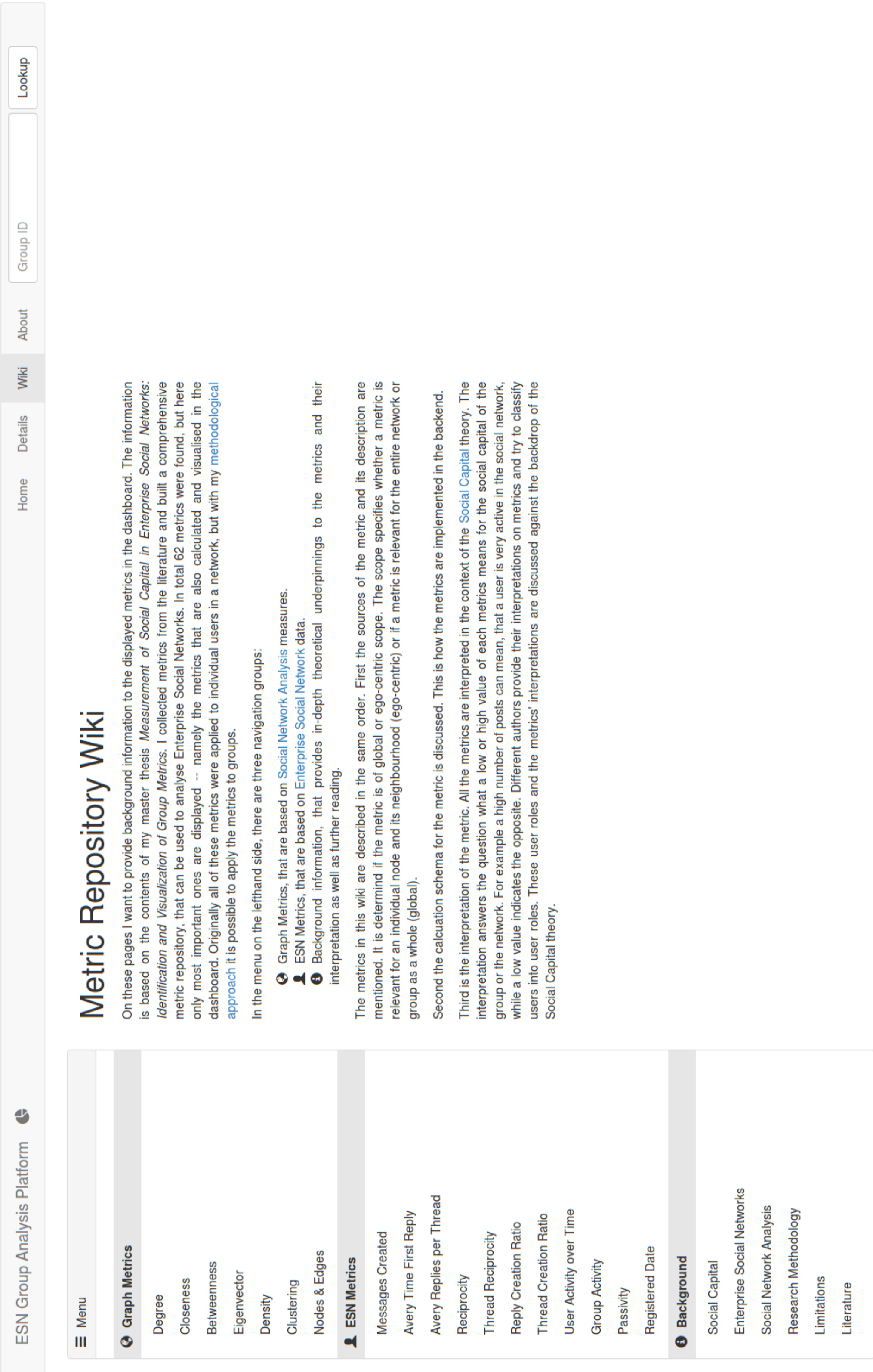


Calculation History

Fri, 03 Feb 2017 11:53:13 GMT
Fri, 03 Feb 2017 11:53:13 GMT
Thu, 02 Feb 2017 16:48:49 GMT
Thu, 02 Feb 2017 16:48:49 GMT

D.3 Wiki

Screenshot of the wiki page.



References

Abela, A. 2013. *No Advanced Presentations by Design: Creating Communication that Drives Action*, San Francisco: Pfeiffer.

Adler, P. S., and Kwon, S.-W. 2002. "SOCIAL CAPITAL: PROSPECTS FOR A NEW CONCEPT." *Academy of Management Review* (27:1), pp. 17–40.

Ali-Hassan, H., Nevo, D., and Wade, M. 2015. "Linking dimensions of social media use to job performance: The role of social capital," *The Journal of Strategic Information Systems* (24:2), pp. 65–89.

Allen, J., James, A. D., and Gamlen, P. 2007. "Formal versus informal knowledge networks in R & D : a case study using social network analysis," *R&D Management* (37:3), pp. 179–196.

Angeletou, S., Rowe, M., and Alani, H. 2011. "Modelling and analysis of user behaviour in online communities," in *International SemanticWeb Conference (ISWC 2011)*, vol. 10, pp. 35–50.

Bächle, M. 2006. "Social software," *Informatik-Spektrum* (29:2), pp. 121–124.

Baker, W. E. 1990. "Market Networks and Corporate Behavior," *American Journal of Sociology* (96:3), pp. 589–625.

Beauchamp, M. A. 1965. "An improved index of centrality," *Behavioral Science* (10:2), pp. 161–163.

Bechmann, A., and Lomborg, S. 2012. "Mapping actor roles in social media: Different perspectives on value creation in theories of user participation," *new media & society* (15:5), pp. 765–781.

Behrendt, S., Richter, A., and Riemer, K. 2014. "Conceptualisation of Digital Traces for the Identification of Informal Networks in Enterprise Social Networks," *ACIS 2014 Proceedings* .

Berger, K., Klier, J., Klier, M., and Richter, A. 2014. "Who is Key...? Characterizing Value Adding Users in Enterprise Social Networks," *22nd European Conference on Information Systems* (Track 21:September).

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

Bonacich, P. 1987. "Power and Centrality : A Family of Measures," *American Journal of Sociology* (92:5), pp. 1170–1182.

Borgatti, S. P., Jones, C., and Everett, M. G. 1998. "Network Measures of Social Capital," *Connections* (21:2), pp. 27–36.

Boyd, D. m., and Ellison, N. B. 2007. "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication* (13:1), pp. 210–230.

Brinton, W. C. 1914. *Graphic methods for presenting facts*, New York, New York, USA: The Engineering Magazine Company, 1st ed.

Brown, J. S., and Duguid, P. 2001. "Structure and Spontaneity knowledge and organization," in *Managing Industrial Knowledge: Creation, Transfer and Utilization*, , London, England: SAGE Publications Ltd, pp. 44–67.

Brown, J. S., and Duguid, P. 2002. *The Social Life of Information*, Boston: Harvard Business School Press.

Burke, M., Kraut, R., and Marlow, C. 2011. "Social Capital on Facebook: Differentiating Uses and Users," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 571–580.

Burt 1992. *Structural holes: The social structure of competition*, Harvard University Press.

Burt, R. 1993. "The Social Structure of Competition," in *Explorations in economic sociology*, vol. 65, , New York: Russel Sage Foundation, pp. 57–91.

Burt, R. 2001. "Structural holes versus network closure as social capital," .

Burt, R. S. 2000. "The Network Structure Of Social Capital," *Research in Organizational Behavior* (22), pp. 345–423.

Clark, H. H., and Brennan, S. E. 1991. "Grounding in communication." in *Perspectives on socially shared cognition.*, B. Lauren, (ed.), Washington: American Psychological Association, pp. 127–149.

Codd, E. F. 1970. "A relational model of data for large shared data banks," *Communications of the ACM* (13:6), pp. 377–387.

Coleman, J. S. 1988. "Social Capital in the Creation of Human Capital," *American Journal of Sociology* (94:6), pp. S95–S120.

Coleman, J. S. 1990. *Foundations of Social Theory*, Harvard University Press.

Cross, R., and Prusak, L. 2002. “The People Who Make Organizations Go - or Stop effectiveness,” *Harvard Business Review* (80:6), pp. 104–112.

DiMicco, J. M., Geyer, W., Millen, D. R., Dugan, C., and Brownholtz, B. 2009. “People sensemaking and relationship building on an enterprise social network site,” *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences* pp. 1–10.

DiMicco, J. M., and Millen, D. R. 2007. “Identity Management : Multiple Presentations of Self in Facebook,” *Group’07* pp. 1–4.

Ellison, N. B., and Boyd, D. M. 2013. “Sociality through social network sites,” *The Oxford Handbook of Internet Studies* pp. 151–172.

Ellison, N. B., Gibbs, J. L., and Weber, M. S. 2015. “The use of enterprise social network sites for knowledge sharing in distributed organizations: The role of organizational affordances,” *American Behavioral Scientist* (59:1), pp. 103–123.

Ellison, N. B., Steinfield, C., and Lampe, C. 2007. “The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites,” *Journal of Computer-Mediated Communication* (12:4), pp. 1143–1168.

Few, S. 2006. *Information Dashboard Design. The Effective Visual Communication of Data*, Sebastopol, CA: O’Reilly Media.

Fischbach, K., Schoder, D., and Gloor, P. 2008. “Analyse informeller Kommunikationsnetzwerke am Beispiel einer Fallstudie,” *WIRTSCHAFTSINFORMATIK* (51:2), pp. 164–174.

Freeman, L. C. 1977. “A Set of Measures of Centrality Based on Betweenness,” *Sociometry* (40:1), pp. 35–41.

Freeman, L. C. 1979. “Centrality in Social Networks: Conceptual Clarification,” *Social Networks* (1), pp. 215–239.

Fulk, J., and Yuan, Y. C. 2013. “Location, motivation, and social capitalization via enterprise social networking,” *Journal of Computer-Mediated Communication* (19:1), pp. 20–37.

Gantz, J., and Reinsel, D. 2011. “Extracting Value from Chaos State of the Universe : An Executive Summary,” *IDC iView* (iView:June), pp. 1–12.

Granovetter, M. S. 1973. “The Strength of Weak Ties,” *American Journal of Sociology* (78:6), pp. 1360–1380.

Granovetter, M. S. 1992. "Problems of Explanation in Economic Sociology," in *Networks and Organizations: Structure, Form, and Action1*, N. Nohria and R. Eccles, (eds.), Boston: HBS Press, pp. 25–56.

Hacker, J., Bernsmann, R., and Riemer, K. 2015. "Dimensions of User Behaviour in Enterprise Social Networks," *Social Knowledge Management in Action* .

Hacker, J., Riemer, K., and Bernsmann, R. 2016. "Discovering Knowledge Actor Roles in Enterprise Social Networks - a Case Study in a Professional Services Firm," *Proceedings of the 37th International Conference on Information Systems* pp. 1–16.

Hansen, D. L., Shneiderman, B., and Smith, M. 2010. "Visualizing threaded conversation networks: Mining message boards and email lists for actionable insights," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (6335 LNCS), pp. 47–62.

Hansen, M. T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Administrative Science Quarterly* (44:1), p. 82.

Hatch, M. J. 2012. "Organizational Culture," in *Organization Theory: Modern, Symbolic, and Postmodern Perspectives*, M. J. Hatch and A. L. Cunliffe, (eds.), Canada: Oxford University Press, pp. 159–173.

Hichert, R., and Gerths, H. 2011. *Geschäftsdiagramme mit Excel nach den SUCCESS-Regeln gestalten*, Freiburg: Haufe-Lexware GmbH & Co KG, 1st ed.

Holtzblatt, L., Drury, Jill, L., Weiss, D., Damianos, L. E., and Cuomo, D. 2013. "Social Media for Organizations Social Media Platform," *Social Media for Organizations* (1:1), pp. 1–21.

Jansen, D. 2002. "Netzwerkansätze in der Organisationsforschung," *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 42* pp. 88–118.

Jarrahi, M. H., and Sawyer, S. 2013. "Social Technologies, Informal Knowledge Practices, and the Enterprise," *Journal of Organizational Computing and Electronic Commerce* (23:1-2), pp. 110–137.

Kaisler, S., Armour, F., Espinosa, J. A., and Money, W. 2013. "Big Data: Issues and Challenges Moving Forward," *2013 46th Hawaii International Conference on System Sciences* pp. 995–1004.

Karoui, M., Dudezert, A., and Leidner, D. E. 2015. "Strategies and symbolism in the adoption of organizational social networking systems," *The Journal of Strategic Information Systems* (24:1), pp. 15–32.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., and Silvestre, B. S. 2011. “Social media? Get serious! Understanding the functional building blocks of social media,” *Business Horizons* (54:3), pp. 241–251.

Kline, J., and Alex-Brown, K. 2013. “The Social Body of Knowledge: Nurturing Organizational Social Capital via Social Media Based Communities of Practice,” *Technical Communication* (60:4), pp. 279–292.

Koka, B. R., and Prescott, J. E. 2002. “Strategic alliances as social capital: A multidimensional view,” *Strategic Management Journal* (23:9), pp. 795–816.

Krackhardt, D. 1992. “The strength of strong ties: The importance of Philos in organizations,” .

Kraut, R. E., Fussell, S. R., Brennan, S. E., and Siegal, J. 2002. “Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work,” in *Distributed work*, P. J. Hinds and S. Kiesler, (eds.), Cambridge, MA: MIT Press, pp. 137–162.

Kuegler, M., Smolnik, S., and Kane, G. 2015. “What’s in IT for employees? Understanding the relationship between use and performance in enterprise social software,” *The Journal of Strategic Information Systems* (24:2), pp. 90–112.

Leidner, D., Koch, H., and Gonzalez, E. 2010. “Assimilating Generation Y IT New Hires into USAAs workforce: The role of an Enterprise 2.0 System,” *MIS Quarterly Executive* (9:4), pp. 229–242.

Leonardi, P. M., Huysman, M., and Steinfield, C. 2013. “Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations,” *Journal of Computer-Mediated Communication* (19:1), pp. 1–19.

Leonardi, P. M., and Meyer, S. R. 2015. “Social Media as Social Lubricant: How Ambient Awareness Eases Knowledge Transfer.” *American Behavioral Scientist* (59:1), pp. 10–34.

Mäntymäki, M., and Riemer, K. 2016. “Enterprise social networking: A knowledge management perspective,” *International Journal of Information Management* (36:6), pp. 1042–1052.

McAfee, A. P. 2006. “Enterprise 2.0: The Dawn of Emergent Collaboration,” *MIT Sloan Management Review* (47:3), pp. 21–28.

McClard, A., and Anderson, K. 2008. “Focus on Facebook: Who Are We Anyway?” *Anthropology News* (49:3), pp. 10–12.

Miller, G. A. 1956. “The Magical Number Seven, Plus or Minus Two Some Limits on Our Capacity for Processing Information,” *Psychological Review* (101:2), pp. 343–352.

Minto, B. 2003. *The Minto Pyramid Principle. Login in Writing, Thinking and Problem Solving*, London, England: Minto International, Inc.

Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., and Guy, I. 2012. “Diversity Among Enterprise Online Communities,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* pp. 2815–2824.

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

Nardi, B. A., and Whittaker, S. 2002. “The Place of Face-to-Face Communication in Distributed Work,” in *Distributed work*, P. Hinds and S. Kiesler, (eds.), Cambridge: MIT Press, pp. 83–113.

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

Portes, A. 1998. “Social capital: Its origins and applications in modern sociology,” *Annual Review of Sociology* (24:1), pp. 1–24.

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

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

Rasiel, E. M. 1999. *The McKinsey Way*, New York: McGraw-Hill.

Razmerita, L., Kirchner, K., and Nabeth, T. 2014. “Social Media in Organizations: Leveraging Personal and Collective Knowledge Processes,” *Journal of Organizational Computing and Electronic Commerce* (24:1), pp. 74–93.

Reagans, R., and Zuckerman, E. W. 2001. “Networks , Diversity , and Productivity : The Social Capital of Corporate R & D Teams,” *Organization Science* (12:4), pp. 502–517.

Richter, A., and Riemer, K. 2009. “Corporate social networking sites – Modes of use and appropriation through co-evolution,” *Acis 2009* (:Schooley 2005), pp. 1–10.

Richter, A., and Riemer, K. 2013. “Malleable End-user Software The authors Malleable End-user Software,” *Business & Information Systems Engineering* (3:1).

Richter, D., Riemer, K., and vom Brocke, J. 2011. “Internet Social Networking,” *Business & Information Systems Engineering* (3:2), pp. 89–101.

Riemer, K. 2005. *Sozialkapital und Kooperation: Zur Rolle von Sozialkapital im Management zwischenbetrieblicher Kooperationsbeziehungen*, Mohr Siebeck, 3rd ed.

Riemer, K., Finke, J., and Hovorka, D. 2015. “Bridging or Bonding : Do Individuals gain Social Capital from Participation in Enterprise Social Networks ?” *Thirty Sixth International Conference on Information Systems* pp. 1–20.

Riemer, K., Scifleet, P., and Reddig, R. 2012. “Powercrowd: Enterprise Social Networking in Professional Service Work: A Case Study of Yammer at Deloitte Australia,” *Working papers - Business Information Systems (BIS)* (WP2012 02:16-May-2012).

Rowe, M., Fernandez, M., Angeletou, S., and Alani, H. 2013. “Community analysis through semantic rules and role composition derivation,” *Journal of Web Semantics* (18), pp. 31–47.

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

Sandefur, R. L., and Laumann, E. O. 1998. “A PARADIGM FOR SOCIAL CAPITAL,” *Rationality and Society* (10:4), pp. 481–501.

Scott, J. 2012. *What is Social Network Analysis?*, Bloomsbury Publishing, 1st ed.

Scott, J., and Carrington, P. 2014. *The SAGE Handbook of Social Network Analysis*, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.

Seibert, S. S. E., Kraimer, M. M. L., and Liden, R. R. C. 2001. “A social capital theory of career success,” *Academy of Management Journal* (44:2), pp. 219–237.

Shneiderman, B. 1996. “The Eyes Have It : A Task by Data Type Taxonomy for Information Visualizations,” *IEEE Symposium on Visual Languages Proceedings* p. 9.

Singh, P. V., Tan, Y., and Mookerjee, V. 2011. “Network Effects: the Influence of Structural Capital on Open Source Project Success,” *Mis Quarterly* (35:4), pp. 813–829.

Smith, M., Hansen, D. L., and Gleave, E. 2009. “Analyzing Enterprise Social Media Networks,” *2009 International Conference on Computational Science and Engineering* (4), pp. 705–710.

Steinfield, C., DiMicco, J. M., Ellison, N. B., and Lampe, C. 2009. “Bowling Online : Social Networking and Social Capital within the Organization,” *Distribution* pp. 245–254.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. 2014. “Socialmedia analytics,” *Business and Information Systems Engineering* (6:2), pp. 89–96.

Treem, J. W., and Leonardi, P. M. 2012. “Social Media use in organizations,” *Communication Yearbook* pp. 143–189.

Tufte, E. 2001. *The Visual Display of Quantitative Information*, Cheshire: CT: Graphics Press, 2nd ed.

Tymon, W. G., and Stumpf, S. a. 2003. “Social capital in the success of knowledge workers,” *Career Development International* (8:1), pp. 12–20.

van Dijk, F. 1997. *Social Ties and Economic Performance*, Boston: Kluwer Academic.

Vie, S. 2008. “Digital Divide 2.0: "Generation M" and Online Social Networking Sites in the Composition Classroom,” *Computers and Composition* (25:1), pp. 9–23.

Viégas, F. B. 2004. “Visualizing the Activity of Individuals in Conversational Cyberspaces,” *Source* (00:C), pp. 1–10.

Viol, J., and Hess, J. 2016. “Information Systems Research on Enterprise Social Networks – A State-of-the-Art Analysis,” *Multikonferenz Wirtschaftsinformatik (MKWI) 2016 - Band I* (0), pp. 351–362.

Ware, C. 2004. *Information Visualization*, San Francisco: Elsevier Inc., 2nd ed.

Wasserman, S., and Faust, K. 1994. *Social network analysis: Methods and applications*, Cambridge University Press.

Watson, G., and Papamarcos, S. 2002. “Social Capital and Organizational Commitment,” *Journal of Business and Psychology* (16:4), pp. 537–552.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., and Faraj, S. 2007. “Information Technology and the Changing Fabric of Organization,” *Organization Science* (18:5), pp. 749–762.

Zhang, J., Qu, Y., Cody, J., and Wu, Y. 2010. “A Case Study of Micro-blogging in the Enterprise: Use, Value, and Related Issues,” *Proceedings of the SIGCHI Conference 2010* pp. 123–132.

Declaration of Authorship

I hereby declare that, to the best of my knowledge and belief, this Master Thesis titled "Measurement of Social Capital in Enterprise Social Networks: Identification and Visualisation of Group Metrics" is my own work. I confirm that each significant contribution to and quotation in this thesis that originates from the work or works of others is indicated by proper use of citation and references.

1st March 2017

Joschka Andreas Hüllmann