Social network analysis in the enterprise: challenges and opportunities

Valentin Burger, David Hock, Ingo Scholtes, Tobias Hoßfeld, David Garcia, Michael Seufert

Angaben zur Veröffentlichung / Publication details:

Burger, Valentin, David Hock, Ingo Scholtes, Tobias Hoßfeld, David Garcia, and Michael Seufert. 2014. "Social network analysis in the enterprise: challenges and opportunities." In *Socioinformatics - the social impact of interactions between humans and IT*, edited by Katharina Zweig, Wolfgang Neuser, Volkmar Pipek, Markus Rohde, and Ingo Scholtes, 95–120. Cham: Springer. https://doi.org/10.1007/978-3-319-09378-9_7.

Social Network Analysis in the Enterprise: Challenges and Opportunities

Valentin Burger, David Hock, Ingo Scholtes, Tobias Hossfeld, David Garcia,
Michael Seufert

Abstract Enterprise social software tools are increasingly being used to support the communication and collaboration between employees, as well as to facilitate the collaborative organisation of information and knowledge within companies. Not only do these tools help to develop and maintain an efficient social organisation, they also produce massive amounts of fine-grained data on collaborations, communication and other forms of social relationships within an enterprise. In this chapter, we argue that the availability of these data provides unique opportunities to monitor and analyse social structures and their impact on the success and performance of individuals, teams, communities and organisations. We further review methods from the planning, design and optimisation of telecommunication networks and discuss challenges arising when wanting to apply them to optimise the structure of enterprise social networks.

1 Introduction

We are currently witnessing a rapidly increasing adoption of technical systems in numerous aspects of everyday life. In particular, the widespread use of information and communication technologies in which interactions and collaborations between humans are an integral part has led to the rise of so-called *socio-technical* systems. A defining characteristic of these systems is that they consists of interwoven social and technical layers, which both are crucial for their functioning. Many examples for such socio-technical systems, like, e.g., social media platforms, collaborative

Valentin Burger, Tobias Hossfeld, Michael Seufert and David Hock
University of Würzburg, Institute of Computer Science, Würzburg, Germany
e-mail: {valentin.burger, hossfeld, seufert, hock}@informatik.uni-wuerzburg.de

Ingo Scholtes and David Garcia
ETH Zürich, Chair of Systems Design, CH-8092 Zürich, Switzerland
e-mail: {ischoltes, dgarcia}@ethz.ch

web applications have recently gained popularity. The popularity and success of these platforms has resulted in the adoption of similar technologies in an enterprise context. Specifically, *enterprise social software tools* are increasingly being used to support the collaboration between employees, as well as to facilitate the collaborative organisation of information and knowledge. Notable examples include groupware systems, collaborative information spaces like Wikis or Blogs, instant messaging, project and knowledge management platforms and - increasingly - social networking services specialised for an enterprise context. While these systems serve different purposes, they have in common that they generate massive amounts of fine-grained data on collaborations, communication and other forms of social relationships between employees. On the one hand, the availability of such data introduces severe privacy issues and thus raises a number of ethical challenges that urgently need to be addressed. On the other hand, such data provide interesting opportunities to gain insights into the structure and dynamics of the social organisation of an enterprise. Not only can important individuals be identified that otherwise may go unnoticed, a monitoring of evolving social structures by means of quantitative measures may also help to identify problems and take adequate countermeasures. A study of quantitative performance indicators - which are often available in an enterprise context - can furthermore provide unique insights into the effect of social structures on the success and performance of individuals, groups and projects.

In this chapter, we argue that the monitoring and optimisation of enterprise social networks provide interesting perspectives for a social informatics research agenda. Intuitively, one could argue that an optimisation of social networks is not easily possible, since they emerge in a self-organised way and thus cannot be influenced or designed. While this is true in many social systems, knowledge from the *planning, design and optimisation of telecommunication networks* can nevertheless be used to analyse the efficiency and resilience of social networks. Furthermore we argue that - through a targeted structuring of teams, the introduction and configuration of communication and collaboration tools as well as the design of corporate policies - the evolution of social networks in an enterprise can - at least to a certain extent - be influenced and shaped. Knowledge from network design may thus be utilised in emerging social organisations to improve resilience and to optimise their efficiency. Similarly, decisions in the design of socio-technical systems can influence the structure of social organisations into which they are embedded.

This chapter is structured as follows: In Section 2 we provide an overview of measures used in the analysis of complex networks, and interpret their meaning in the context of enterprise social networks. In Section 3, we show how one can use these measures to monitor the structure and evolution of social networks extracted from socio-technical systems. Section 4 introduces basic notions used in the optimisation of communication networks and discusses how these approaches can be used in the context of social networks. We further highlight research challenges arising in the modelling, analysis and optimisation of enterprise social networks. Highlighting links between research in the fields of *network planning and design* and *social informatics* - which are currently not well integrated - we conclude in section 5.

2 Quantitative Analysis of Complex Networks

The increasing availability of data that allows to reconstruct networks of interactions between elements in complex systems has led to a massive surge of interest in the quantitative analysis of complex networks. During the last few decades, a comprehensive set of measures has been introduced, which allow to quantify characteristics of complex networks. Referring to available reference books for more details [10, 17], in the following, we provide a brief overview of these measures and interpret their meaning in the context of enterprise social networks. We particularly categorise measures into *node-centric* measures, which are targeted at capturing characteristics of individual nodes, as well as *network-centric* measures, which capture systemic properties of complex networks. In the following, we refer to a network $G=(V,E)$, which consists of a set V of *nodes* as well as a set $E \subseteq V \times V$ of links that interconnect nodes. In the context of enterprise social networks, we commonly assume that nodes represent *employees* or *co-workers* within an enterprise, while links between them are thought to represent some form of *social interaction*, like, e.g., an exchange of information, a conversation across E-Mail, instant messaging or voice communication services, or the collaboration in the context of a particular project. The ability to automatically reconstruct enterprise social networks require data on these interactions to be recorded, which typically implies that they are mediated via some type of technical system. However, the increasing adoption of modern wearable computing and sensing technologies highlights scenarios where networks can also be constructed from direct interpersonal communication between employees as well as their mobility traces [11].

For the remainder of this section, unless stated otherwise, we assume that networks are *undirected*, i.e. a link (v,w) between two nodes v and w implies that the reverse link (w,v) exists, in which case both links can be conveniently represented by a single undirected link. Even though the number, frequency or intensity of recorded social interactions can often be used to establish a notion of *link weights*, for the sake of simplicity, we further assume that networks are unweighted, i.e. the weight or strength of all links is the same. A simple example for such an undirected, unweighted network - representing social interactions between members of a software development team - is shown in Figure 1.

2.1 Node-centric Metrics: Centrality and Topological Embedding

A basic task in the analysis of complex networks is to quantify the importance - or centrality - of individual nodes, as well as how they are embedded in the overall topology. In the following, we thus give a brief overview of different measures that have been proposed for this purpose, and how they can be interpreted in the context of enterprise social networks. For their interpretation, it is important to consider the semantics of links and the resulting meaning of the network topology in the given context.

Node degree. A particularly simple measure which is often used to capture the importance of a node is its *degree*, which is defined as the number of direct neighbours to which it is connected. A natural tendency of (social) networks occurring in many contexts is that they exhibit *heavy tailed distributions* of node degrees, implying that there are a few nodes whose degrees are magnitudes larger than the degrees of the majority of nodes in the network. In the context of enterprise social networks where links represent collaborations, an exchange of information or communication, the degree of nodes can be used as the most basic proxy for the popularity or importance of the persons they represent. While heavy-tail degree distributions arise naturally in social networks, they can be used to evaluate the *centralisation* of the social organisation of collaborating teams. Furthermore it has been argued that the cognitive capabilities of humans pose a limit to the number of stable inter-personal social relations [2]. Thus, structures in which most of the links are concentrated on only a few nodes, can indicate situations in which central employees are being overburdened by communication, This can possibly have negative consequences for the efficiency of a social organisation. Furthermore, random networks with heavy-tail degree distributions have a tendency to be vulnerable against the loss of high-degree nodes, meaning that the network can be disconnected even though only a small fraction of its most connected nodes are removed [1]. As such, the degree of centralisation of an enterprise social network in terms of node degrees can be seen as a simple proxy for the resilience of a social organisation against the loss of its most connected members.

Path-based centrality measures. A different set of measures for the importance of nodes in a network are those which are based on the topology of *shortest paths* between nodes in a network. One important example is the *betweenness cen-*

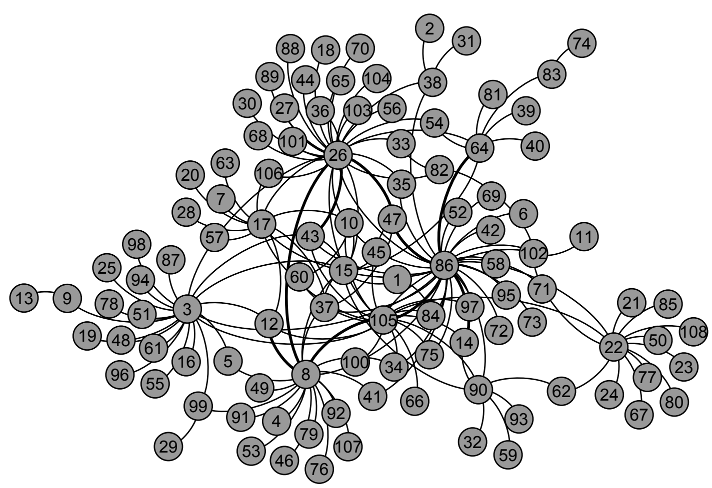


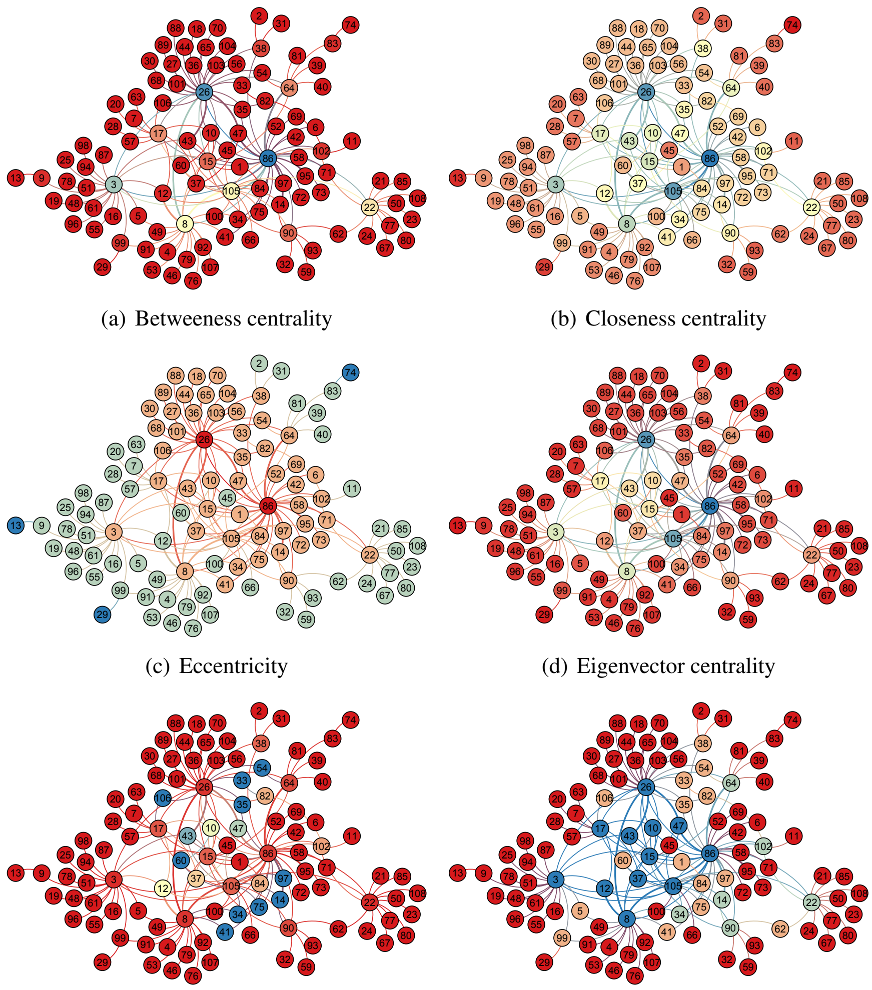
Fig. 1 An example network $G = (V,E)$ consisting of a set of *nodes* V and *links* $E \subseteq V \times V$.

trality of a node v, which is defined as the number (or fraction) of shortest paths between any pair of nodes that pass through node v [4]. Similarly, the *closeness centrality* of a node v is defined based on the average distance of a node v from any other node in the network. To obtain a measure of centrality in which higher values indicate more central nodes, the inverse of the average distance is typically used, meaning that a node with closeness centrality 1 is directly connected to any other node, while its closeness centrality tends to zero, as the average distance to other nodes tends to infinity. Instead of the average distance to all other nodes, one can alternatively study the maximum distance of a node v to any other node in the network, which is called its *eccentricity*. The betweenness and closeness centrality of nodes in an example network, as well as their eccentricity is depicted in Figure 2(a)-2(c).

While nodes with high degree have a tendency to be important also in terms of path-based centrality measures, this correlation does not hold necessarily. Nodes with high degree can still be in the periphery of a network in terms of their average or maximum distance to all other nodes, meaning that they have small eccentricity and closeness centrality. Conversely, nodes with small degree can nevertheless reside at the core of a network through which many paths pass, meaning that they have high betweenness centrality. Path-based centrality measures thus capture a different dimension of topological importance and can thus play an important role in the monitoring, analysis and optimisation of enterprise social networks. In particular, individuals with high betweenness centrality may go rather unnoticed as they are not necessarily in contact with *many* colleagues. Nevertheless, their loss will have considerable impact on the flow of information, as it will change a seizable fraction of shortest paths between other nodes in the network. Furthermore, individuals with high betweenness centrality often play the role of *mediators*, which interconnect different parts of an organisation and bridge information between different communities. At the same time, a highly skewed distribution of betweenness centrality can be interpreted as a sign of high centralisation, which potentially poses a risk for efficiency and resilience of social organisations. While betweenness centrality captures shortest paths *passing through* a node, closeness centrality and eccentricity focus on the length of paths *starting or ending in a node*. As such, they capture how individuals are able to receive and propagate information travelling across shortest paths: Individuals with high closeness centrality can be seen as good information spreaders, since they can propagate information throughout the network most quickly. Nodes with small closeness centrality on the other hand are on in the periphery of a social organisation, thus receiving information - on average - later than others. Similarly, for nodes with high eccentricity there exist other nodes in the network that can only be reached via long paths. Individuals that play a central role in a social organisations should thus - in general - exhibit high closeness centrality and small eccentricity.

Clustering coefficient. Apart from different dimensions of importance introduced above, another important characteristic of nodes is how they are embedded into the topology of a network. One interesting aspect is, for instance, whether the neighbours of node v are also connected to each other, or - in other words - whether *triads* $(v,x),(v,y)$ around a node v are closed. To quantify this property, the *clustering coefficient* of a node v is defined as the fraction of pairs of neighbours x and y of v for which a link (x,y) exists. The clustering coefficient of nodes in a sample communication network is visualised in Figure 2(e). In the context of enterprise social networks, several different interpretations for the presence of nodes with high clustering coefficient are possible: First of all, naturally evolving social networks are known to have a - compared to random networks - high average clustering coefficient. At the same time, they exhibit a small diameter that is due to so-called *weak ties* which bridge the local cluster structures around nodes. Social networks with such a combination of high clustering coefficient and small diameter are usually called *small worlds*. Different from general networks with low diameter, small worlds typically have the property that they are *navigable* for humans, i.e. individuals are able to locally route information along short paths without global knowledge about the network topology. One property that enables individuals to quickly identify neighbours which lie on short paths to a given target is *funneling*, i.e. the fact most short paths pass only through a small set of neighbours which have connections outside local cluster structures. As such, the clustering coefficient of enterprise social networks can be used to quantify aspects that influence their navigability, an important property for the routing of information. Being aware which colleagues represent weak ties to other communities (and which thus transcend local clustering structures) is likely to be important, e.g. in order to quickly identify which colleagues have a particular expertise or work on similar projects, even if they are not directly connected to an individual. Furthermore, a high clustering coefficient of a node can be used as a proxy for the impact of removing this individual from a team: For a node v with high clustering coefficient, most of v's neighbours can still communicate or collaborate with each other even if v is removed from the network. Similarly, a high clustering coefficient can help to mitigate the overload of a central node v, since communication between two neighbours u and w can alternatively bypass v via a direct link (u,w).

Coreness. Another aspect of the embedding of individual nodes into the topology of a network is captured by their *coreness*. The k-core of a network is defined as the largest subgraph of a network, in which each node has a degree of at least k. Based on this decomposition in different k-cores, the *coreness* of a node is the maximum k-core to which it belongs. The coreness of nodes in the example network is shown in Figure 2(f). In particular, the k-core of a network is the largest connected component which is left when repeatedly removing all nodes with degree smaller than k. As such, the k-core decomposition of a network, as well as the coreness of nodes, plays an important role in the analysis of resilience of social network structures against cascading processes. The presence of k-cores with high k in an enterprise social network can be related to its ability to withstand *turnover* of employees, as well as potential cascades or network effects potentially triggered by individuals leaving a company.



(a) Betweeness centrality (b) Closeness centrality

(c) Eccentricity (d) Eigenvector centrality

(e) Clustering coefficient (f) Coreness

Fig. 2 Overview of node-centric metrics in an example network. The score of each node is depicted by its colour (red: low, yellow: medium, blue: high).

2.2 Network-centric Metrics: Resilience and Efficiency

Apart from measures that address the importance and topological embedding of individual nodes, an important further contribution of network theory is the provision of *aggregate, network-centric measures* that can be used to capture *systemic properties* of complex networks. In the following, we briefly introduce a set of network-centric measures that can be related to two particular systemic properties of complex networks: their resilience against failing nodes or links as well as their efficiency in terms of information propagation. We then interpret them in the context of enterprise social networks.

Network size, compactness and average degree. The simplest possible aggregate quantities of a complex network are the number of its nodes and edges. Based on these quantities, the *compactness* of a network can be defined as the ratio between the number of edges and the number of edges that could possibly exist in a network with the same number of nodes. The *average degree* is defined as the average of the degrees of all nodes. For networks with scale-free, heavy-tail degree distributions, the average degree is - in general - not a good representation for the typical degree of connectivity in the system. In such networks the degree of most nodes is in fact much smaller than the average, while a few nodes have degrees orders of magnitude larger than the average degree. In social networks where the average degree is a good representation of the typical degree of connectivity, it allows to analyse which of the individuals have more or less connections than the typical node. The compactness of a network - or of different of its subgraphs - is an interesting measure to evaluate one aspect of the *group cohesiveness* of a social organisation. While social networks with high compactness exhibit a high level of cohesiveness, they are likely to run into scalability issues as the network grows. In general, large-scale social networks which support efficient information exchange are expected to be sparse, meaning that their compactness is relatively small.

Diameter and average distance. As argued in Section 2.1, a further important characteristic of a network topology is its *diameter*, which is defined as the longest shortest path between any two nodes in the network. Similarly, the *average distance* gives the average length of shortest paths between any pair of nodes. Both quantities play an important role in the analysis of enterprise social networks, since they quantify how efficient individuals can communicate across shortest paths. A large diameter indicates the presence of at least one pair of individuals that are connected only via a long path. Even worse, a large average distance indicates that the characteristic length of shortest paths between individuals is large. Enterprise social networks supporting efficient information flow between employees are thus expected to exhibit short average distance and diameter.

Measures of connectivity. Capturing the resilience of a network, its node (or edge) connectivity is defined as the amount of nodes (or links) that have to be removed before it falls apart in different components. Both notions of connectivity are illustrated in the network shown in Figure 3, which has a node connectivity of one and a link connectivity of two (assuming that both subgraphs have higher node and link connectivity). Since each node is connected to a network by at least one link, the link connectivity of a network is always at least as high as its node connectivity. A different approach to quantify the connectivity of a complex network is in terms of its *algebraic connectivity*, a measure which is defined as the second smallest eigenvalue in the spectrum of eigenvalues of a networks' Laplacian matrix [1]. The *algebraic connectivity* can be seen as a generalisation of a network's *connectedness*, where connectedness captures whether all nodes in the network belong to a single connected component. An algebraic connectivity of zero indicates that the network is disconnected, while connected networks exhibit non-zero values. For connected networks, the actual value of algebraic connectivity has been shown to reflect how "well-connected" the network is. In particular, a large algebraic connectivity indicates a) high node and link connectivity, and b) small diameter, while small values indicate the opposite [3, 16, 18]. In the context of enterprise social networks, node, link and algebraic connectivity are important approaches to quantify both their resilience and efficiency. Node and link connectivity is crucial for resilient social organisations, since the failure of low connectivity nodes or links can severely impact the network structure, for example, by a separation of communities. The node and link connectivity of networks can be used to identify such critical nodes and links. Furthermore, networks with node and link connectivity exhibit *small cuts* in the topology, which - apart from being a threat to resilience - can be interpreted as bottlenecks that inhibit the diffusion of information. Combining both node and link connectivity and diameter, algebraic connectivity can be used as a measure which jointly captures the efficiency of information flow in a social organisation as well as its resilience: First of all, a large value of algebraic connectivity indicates that all individuals can communicate with each other via short paths. However, it also shows that there are no bottlenecks in the sense that a large fraction of paths necessarily pass through a small set of nodes or links. Algebraic connectivity can thus be interpreted as a measure for the *cohesiveness* of a social organisation.

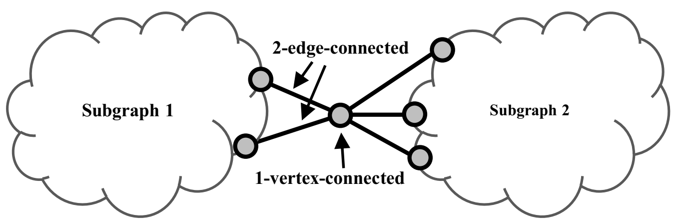


Fig. 3 Illustration of the difference between edge-connectivity and vertex-connectivity.

[1] The Laplacian matrix $\mathbf{L}$ of an undirected network is commonly defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{A}$ is the usual binary adjacency matrix of the network and $\mathbf{D}$ is a a diagonal matrix where diagonal elements contain the degree sequence of the network.

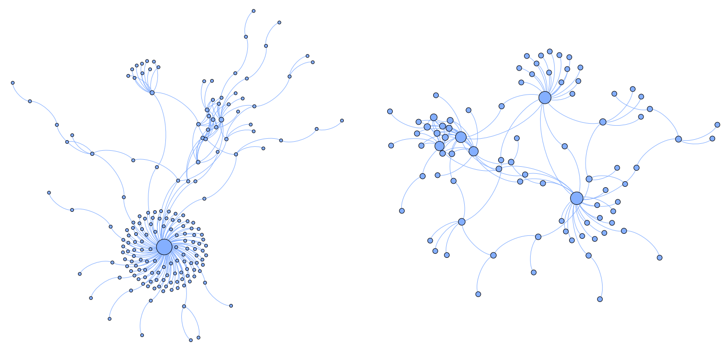
3 Mining Socio-Technical Systems: Application of Network-theoretic Measures

The measures introduced in Section 2.2, as well as their interpretations provided above, highlight interesting opportunities for the monitoring and analysis of enterprise social networks. In practice, data which are suitable to construct and analyse such social networks can come from a variety of social software used in an enterprise context, including social networking tools, collaboration platforms, messaging systems or project management tools. In this section we exemplify this approach using a data set of time-stamped collaborations obtained from a web-based project management tool used by distributed software development teams. In particular - utilising data on Open Source Software communities which have previously been used in the studies [6, 15, 19, 20] - we exemplify some of the metrics introduced in Section 2.2 and provide a complementary, in-depth analysis of the social organisation of two projects that are the GENTOO project and the ECLIPSE project.

3.1 Monitoring Open Source Software Communities

A particularly important and widely used class of enterprise social software that allows to construct social networks are project management tools which support the collaboration, communication or task-allocation in distributed teams. In the context of distributed software development teams, *issue tracking tools* are an important example which allow to report, prioritise and filter reports about software defects, as well as coordinate the efforts to solve them. Such tools are widely used not only within an enterprise scenario, but also in *Open Source Software* (OSS) projects. Since these tools are publicly available to users and contributors of the project, it is possible to extract rich data on the evolving social organisation of these projects. In the following, we thus utilise these data as a proxy to study evolving social structures of humans collaborating on a project. We particularly focus on OSS projects which use BUGZILLA, a popular issue-tracking tool which is widely used in the development of both open source and commercial software projects. While the same data set has been used in [19] to study 14 OSS communities, here we provide detailed results for two major OSS communities: The first is GENTOO a project developing a LINUX-based operating system. The second project is ECLIPSE, which develops and maintains one of the most popular integrated software development environments.

Our approach is based on a construction of *evolving social networks* based on time-stamped interactions between team members that are recorded in the BUGZILLA installation of a project. All recorded interactions within BUGZILLA evolve around *bug reports*, which typically contain a collection of information about a particular software defect. Here, we make use of so-called *Assign* and *CC* interactions, which have a special semantics in the context of issue tracking: A *CC* interaction between a team member *A* and *B* implies that *A* forwards information about a

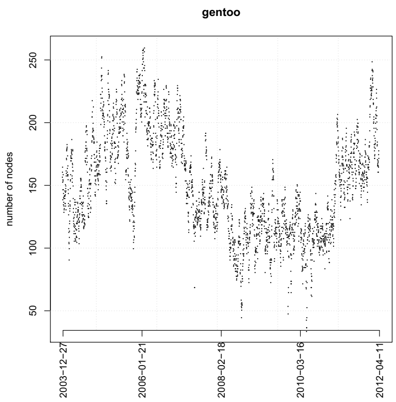


(a) 2007 (b) 2009

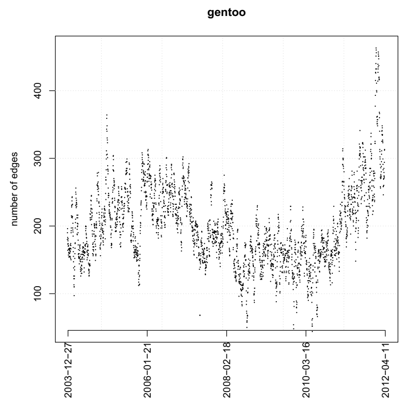
Fig. 4 Two collaboration networks of the GENTOO community, constructed from interactions recorded over a period of 14 days in mid 2007 and mid 2009.

bug report to team member B. An *Assign* interaction between A and B means that A assigns the task of resolving a bug (e.g., by providing a software fix or workaround) to another team member B. In the following we take a maximally simple perspective and say that any interaction between A and B implies that A is aware of B, thus allowing us to construct a network of collaborating team members. In particular, we consciously sacrifice the additional semantics of different interaction types, as well as their potential implications for the role of individuals, for the sake of simplicity. Since all interaction events recorded in BUGZILLA have precise time stamps, we can further construct *time slices* $[t, t+\delta]$ of social networks by only taking into account interactions happening between time stamps t and $t+\delta$. Using a window size δ of 14 days and an increment $t \to t'$ of one day, we perform a *sliding window analysis*, eventually obtaining a sequence of evolving collaboration networks covering the periods $[t, t+\delta], [t', t'+\delta]$ and so forth. Figure 4 shows two example networks constructed from 14 day time slices of the GENTOO project. Nodes in this network represent team members who have been active in the project's BUGZILLA installation within a period of two weeks.

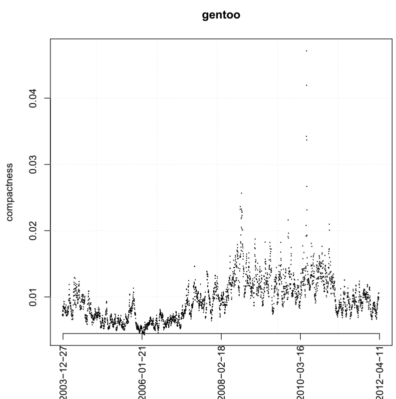
Having constructed sequences of collaboration networks for a project allows to apply network-centric measures, thus capturing characteristics of the project's social organisation. Figure 5 shows the evolution of six metrics introduced in Section 2.2 for the project GENTOO. Figures 5(a) and 5(b) show the number of nodes and links in the largest connected component of the collaboration networks spanning a period of two weeks. One observes significant changes in the number of nodes and edges, highlighting two remarkable phases of growth between December 2003 and Febru-



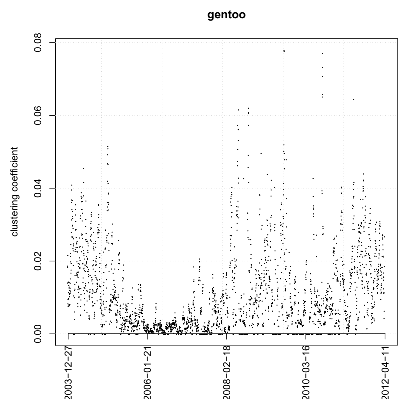
(a) Number of nodes



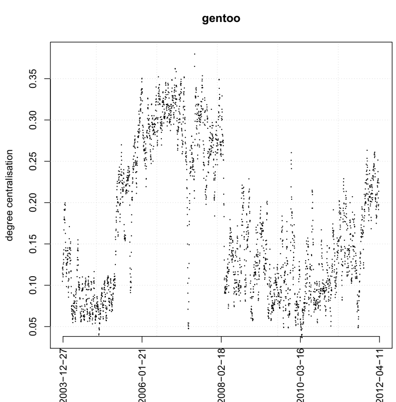
(b) Number of edges



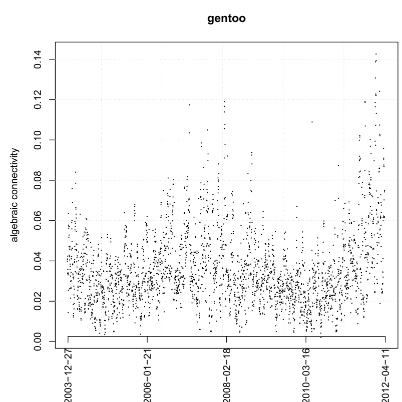
(c) Compactness



(d) Average clustering coefficient



(e) Degree centralisation



(f) Algebraic connectivity

Fig. 5 Evolution of network measures capturing social organisation in the GENTOO project.

ary 2006, as well as between June 2010 and April 2012, when our data collection stopped. In addition, a phase during which the number of nodes decreased can be observed between January 2006 and June 2008, followed by a phase of stagnation between June 2008 and June 2010. The number of links representing interactions between team members qualitatively follows the dynamics in the number of active nodes. The number of nodes and links highlights a non-stationary level of team activity and can thus help to interpret the dynamics of other characteristics that are typically affected by the network size. As argued in Section 2.2, the compactness of a network is a simple size-independent measures which can be interpreted as a particularly simple proxy for the *cohesiveness* of a social organisation. In the GENTOO community, we observe a first phase of decreasing compactness between December 2003 and February 2006, which coincides with the first phase of growth. After a phase of stagnation and moderate increase between February 2006 and March 2008, the compactness of the social network doubled around March 2008, indicating an increase in cohesion. The average clustering coefficient shown in Figure 5(d) shows a similar dynamics. A first phase lasting until February 2006 shows a remarkable decrease of the average clustering coefficient. During a second phase between February 2006 and March 2008, the average clustering coefficient is remarkably small. The increasing compactness starting in June 2008 was accompanied by an increasing embedding of nodes in densely connected clusters. In Figure 5(d), the phase between February 2006 and March 2008 is particularly noteworthy. One can get a clearer picture of the processes shaping the social organisation during this phase by considering additional network-centric metrics. Figure 5(e) shows the evolution of degree centralisation, a measure defined based on the distribution of node degrees. A value of one represents a maximally centralised situation in which all nodes are only connected to a single central node, while a value of zero represents a situation where all degrees of nodes are equal. The degree centralisation shows a remarkable dynamics, exhibiting a highly centralised phased between mid 2005 and March 2008, with centralisation quickly dropping around March 2008. An interview with past and current members of the GENTOO issue tracking team performed in [15] revealed that - between mid 2005 and March 2008 - most of the work associated with the processing of bug reports was done by a *single team member*. Following a dispute with other team members, and being overburdened with tasks, this central member left the project unexpectedly in March 2008. Following this event, the community actively took efforts to reorganise the bug handling process, which is likely to be the reason behind the increasing compactness and clustering coefficient. The evolution of algebraic connectivity depicted in Figure 5(f) shows a high variability, with a slightly increasing trend between the end of 2005 and the beginning of 2008. Interestingly, the reorganisation of the community following the loss of the central contributor was accompanied by an observable decrease of algebraic connectivity until mid 2010, after which it increased significantly.

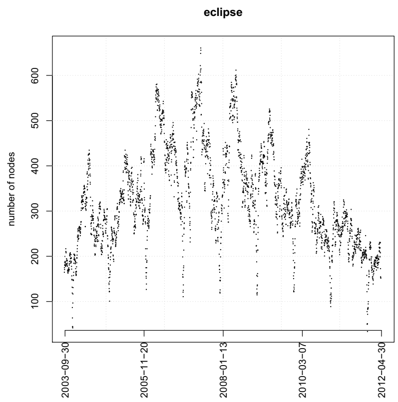
Two collaboration networks illustrating the difference in social organisation during the presence of the central contributor between 2005 and 2008, compared to the time after she left are shown in Figure 4(a) and 4(b). It is tempting to relate the obvious changes in the social organisation discussed above with changes in the performance of the bug handling process during the same period. A study of bug handling performance in the GENTOO community has recently been presented in [15]. It shows that the performance in terms of number of reported/resolved bugs, as well as in terms of the time taken between the submission of a bug report and the first response of a community member, show an interesting dynamics that is likely to be correlated with the evolution of social organisation. In particular, here it was shown the performance of the GENTOO bug handling community generally increased until early 2008. A rapid increase in the response time as well as in the number of open bug reports can be observed at the time when the central contributor left, followed by a phase of stagnation until early 2011 after which performance increased again.

Applying the same measures as above and highlighting differences in the dynamics of social organisation, we now turn our attention to the project ECLIPSE. Figure 6 shows the evolution of six network-centric measures over a period of almost ten years. A first remarkable observation is a pronounced periodicity in the number of nodes and edges, as well as in compactness and the average clustering coefficient. Both the number of nodes and edges in the collaboration network experience a steep increase of up to 500 % roughly once a year. While we cannot make definite statements about the underlying reasons, it is likely that this periodicity is related to the project's release cycle, which aims at one release per year. This increase in activity is associated with increases in both the compactness (Figure 6(c)) as well as the clustering coefficient (Figure 6(d)). A further remarkable fact is that - while slight periodic peaks can be observed - compared to the GENTOO project - degree centralisation remains at a rather moderate level also in phases of high activity. One may interpret this as a sign of a *healthy* social organisation, in the sense that an increase of activity is associated with an increase of cohesion, rather than an unproportionate burdening of a few team members.

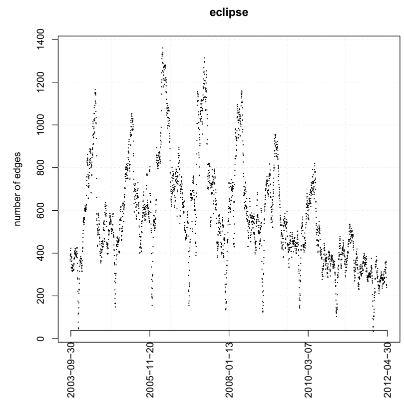
3.2 Analysing Resilience in Online Social Networks

Online social networks are socio-technical systems in which users interact through an online medium, overcoming some of the limitations of verbal face-to-face communication. To improve user experience, the technological component of an online social network is subject to be changed and redesigned, introducing modifications in the fundamental way in which individuals communicate. The impact of such changes in a social system is not trivial, as user reactions are coupled to each other. A technological change, such as a new user interface, can trigger some users to leave the social network, which can decrease the quality of the experience of its friends. This mechanism leads to cascades that can potentially lead large amounts of users away from the social network.

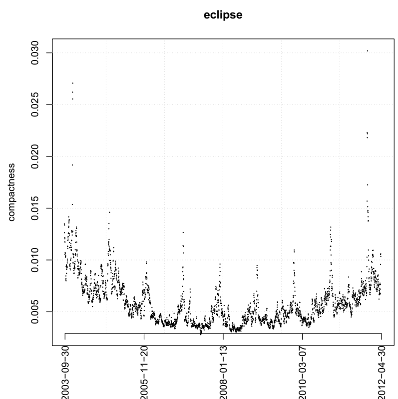
Users leaving a social network can be modelled through a decision process, in which a user receives a benefit and a cost associated with being active in the network. In terms of social interaction, the benefit perceived by a user comes from its contacts, either in the form of information, or as attention. The benefit of a user



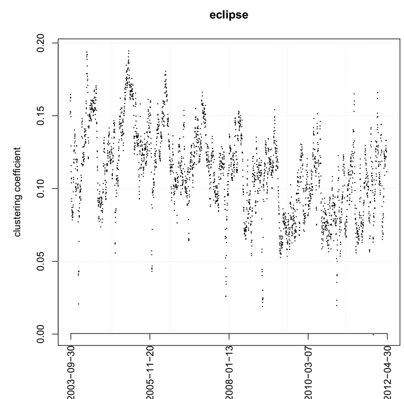
(a) Number of nodes



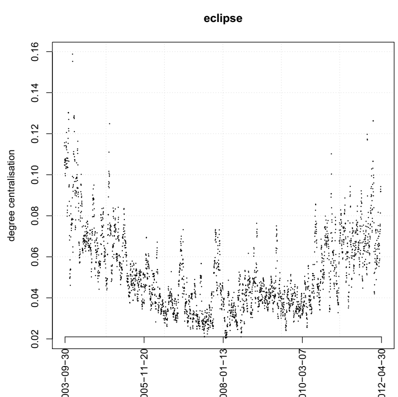
(b) Number of edges



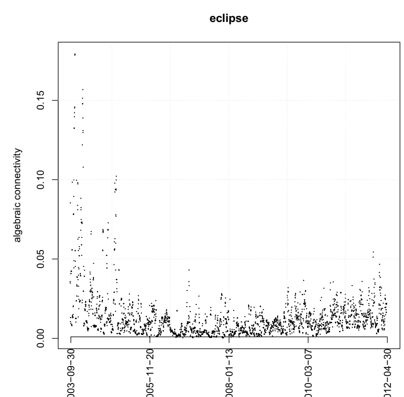
(c) Compactness



(d) Average clustering coefficient



(e) Degree centralisation



(f) Algebraic connectivity

Fig. 6 Evolution of network measures capturing social organisation in the ECLIPSE project.

increases with the amount of active friends, and decrease every time a friend becomes inactive. Costs do not necessarily need to be economic, they can also include the time spent in a social network, or the opportunities lost by not using other platforms. This cost can be increased due to changes in the user interface, or due to service limitations, threatening the cohesion of the network as a society. This model allowed us to analyse the cascades of users leaving the network [5], which stop in subsets of the network that corresponds to the *k*-cores explained in Section 2.1. Thus, the *k*-core decomposition of a social network allows us to measure its social resilience, i.e. how does the network withstand external shocks and stresses.

We empirically analysed the social resilience of a variety of online social networks, through their k-core decomposition. Such empirical analysis, including large sites like Facebook, MySpace, and Friendster, showed that the topology of these networks can vary a lot in terms of resilience, calling for methods that can increase this desired property. For example, friendship links can be recommended in a way such that links that increase coreness are encouraged, or changes can be introduced gradually to limit cascades of departing users.

These results show that a quantitative analysis of data on enterprise social networks can provide interesting insights into the evolving social organisation of teams, projects or communities. In the case of the GENTOO project, our results show that a monitoring of degree centralisation and average clustering coefficient may have been used as an early indicator for a detrimental evolution of social structures. Furthermore, it is at least conceivable that a targeted optimisation of the network's resilience against the loss of its most central node may have prevented downstream problems with the performance of the bug handling process. In the following chapter, we thus review approaches from the planning and design of telecommunication networks, and discuss their possible application to enterprise social networks.

4 Network Planning and Design: Application to Enterprise Social Networks

Analogously to enterprise social networks, the overload or loss of nodes can severely impact telecommunication networks. Therefore, telecommunication providers monitor their networks to allow for an early identification of emerging problems and to take appropriate measures to mitigate their impact on performance. The most important goals of these interventions are to optimise resilience of the network against failing nodes or links, but also to balance load across links and nodes in order to avoid congestion and overload of single nodes which may significantly decrease the performance and efficiency of the entire network – similar to the GENTOO project. Typical interventions include the addition of nodes or links to increase resilience, the rewiring of links or an adaptation of *link capacities* to optimise traffic flows, or the addition of special functionality nodes to manage or monitor regions of a network. In more general terms, *network planning and design* refers to the process of designing the topology of telecommunication networks in a way that optimises some notion of *value*, while keeping *costs* as small as possible. In the following, we will sketch how network optimisation is achieved in general in the context of telecommunication networks. We then we apply it exemplarily to a collaboration network extracted from the GENTOO Open Source community and summarise research challenges arising when wanting to optimise and influence social network structures.

4.1 Network Optimisation

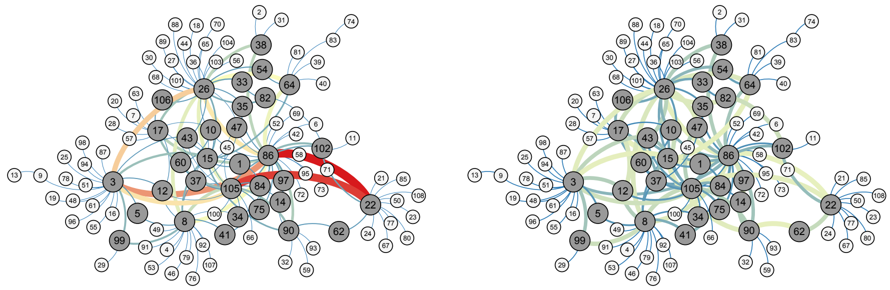
In general, the optimisation of network topologies consists of the following three steps: In a first step, an optimisation objective has to be defined based on a notion of *value* defined for particular network topologies. Depending on the context and the associated objective this value can be defined based on different, not necessarily correlated, typically network-centric measures. To give an example, if resilience is to be optimised, one may utilise the average coreness of nodes or the (algebraic) connectivity of the network, while one may chose the average shortest path length, if the latency of communication as to be optimised. In the context of enterprise social networks, the value could, e.g., be defined based on measures capturing communication efficiency, resilience or aspects which influence work atmosphere, thus seeking to balance different aspects by means of a multi-objective optimisation. In general, in the following we assume that the value of a network can be defined as a function of the network topology, which uses the topological structure, node properties, and link weights to quantify the value of the network in a particular context. As second step, an adequate degree of abstraction has to be found to model the telecommunication network for the optimisation process. There is a wide range of different abstraction levels reaching from simple adjacency and distance matrices over partial or complete lists of all possible routing paths up to object oriented representations of each single link and node. The chosen model is one of the essential influence factors for the thirds and final step: the actual optimisation. Probably most optimisation methods known in science and engineering can and have already been applied to network optimisation, including, e.g., (mixed integer) linear programming approaches, different heuristics such as simple greedy mechanisms, simulated annealing, evolutionary algorithms, and - if computationally feasible - the exhaustive evaluation of the entire search space of optimisation options. Each of these methods has advantages and disadvantages, in particular concerning the optimality of the results, the computational complexity of the optimisation, and the capability to handle large network topologies. But they have in common that they aim at maximising the value of the network while balancing it with the associated costs.

4.2 Application to an OSS Collaboration Network

In section 3, we have seen that a monitoring of network-centric measures capturing resilience (like, e.g., degree centralisation) or aspects that influence communication efficiency (like, e.g., algebraic connectivity) can provide valuable insights into the changing social organisation of software development teams. Furthermore, we have argued above that these aspects are important criteria which are typically accounted for in the design and planning of telecommunication networks. In the following, we briefly describe methods used in network optimisation and discuss their potential applications in the context of enterprise social networks. We further use one snapshot of the collaboration network extracted from the GENTOO community during a period of four weeks in May 2002 to illustrate their application in social networks.

Routing Optimisation. To save operational costs the network infrastructure and resources need to be highly utilised without overloading single entities. To cope with the daily dynamics of traffic, resources need to be added in peak hours. This can be achieved by flexible resource allocation and dynamic routing. Although many mechanisms have been investigated to use network resources efficiently, routes in the Internet are still static in many cases. Especially in case of link or node failure, alternative routes that have to carry the traffic of the broken connections are likely overloaded. The aim of routing optimisation is to balance the load on the links of a network [8]. To add resilience, routing is further optimised such that the load is balanced in case of link or node failure. In companies load balancing is important to unburden central employees. It is also reasonable to route, or assign tasks in such a way that information flow is resilient against node or link outages. That means that the workload is still balanced among employees if worker fails. Assuming that each node produces the same amount of tasks, the load on the communication channels can be estimated. To balance the load on the communication channels we consider the 2-core of the communication network, since there exist no alternative links for stub nodes. Stub nodes are nodes that are connected only with one link to the large component to the graph. Figure 7(a) shows the link load in the OSS network if routing is not optimised. The edge colour depicts the utilisation of a link. If routing is not optimised, the links from 22 to 86 and from 22 to 105 are highly loaded, probably resulting in an overload of the involved individuals. Furthermore, if individual 105 fails all tasks forwarded by node 22 would have to be completed by node 86, which significantly burdens the central contributor. To optimise routing, load is taken from the overloaded links and shifted to alternative routes. In this example load is shifted to node 62. In enterprise social networks, shifting load could be realised by delegating tasks to a different set of workers. Thus, the load is balanced among the paths in the network. If node 105 fails, the tasks originating from node 22 or one of its neighbours, are still shared by node 62 and 86. The load on the links with optimised routing is depicted in Figure 7(b).

Controller Placement. In communication networks controllers are needed for authentication, authorisation and connection establishment, but also for dynamic

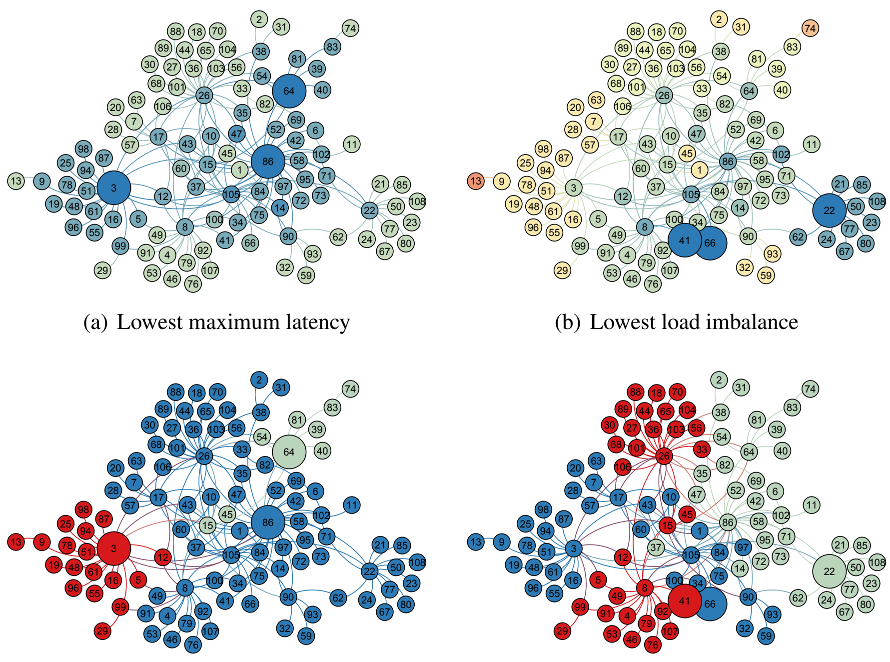


(a) Unoptimised routing (b) Optimised routing

Fig. 7 Illustration of different routing layouts in the GENTOO collaboration network. The color and the strength of the edges depicts the load on the link (red: high, yellow: medium, blue: low).

network control adding network functions and rerouting. To monitor network flows efficiently and to be able to access and control all routers in the network efficiently, it is important to place the controllers strategically. In telecommunication networks this process is called controller placement [7]. “Controller placement” can be applied to companies in the sense that workers are chosen as controllers to efficiently spread information or to promote workers which will take a lead in assigning tasks and delegating responsibilities. This type of “promotion” is actually an important mechanism in the bug handling communities of Open Source Software communities, since it is typically only a small set of privileged individuals which is allowed to assign tasks to other community members or developers. If the GENTOO network is considered, and it is assumed that community has the capacity to promote three workers, the question is which workers to choose. Figure 8(a) shows the collaboration network of the GENTOO project. Three controllers - highlighted by a larger node size - are placed in the network in a way that optimises the maximum latency from each node to the nearest controller. Here, nodes 3, 64 and 86 are selected. Node 86 is the central contributor in the GENTOO community and has direct access to a large part of the community. Nodes 3 and 64 are less important, but nevertheless central, nodes covering different parts of the community. Figure 8(c) shows the nodes that are assigned to each of the three controllers. The number of worker assigned to nodes 3 and 64 is small compared with the much larger amount of workers assigned to node 86. That means node 86 would be responsible for many members in the community, which puts a high load on this central contributor. To unburden this central contributor, responsibilities can be delegated to different members, in a way, that each leader is responsible for an equal amount of members in the community. In Figure 8(d) the controller placement is optimised for lowest load imbalance. Three controllers are placed and associated to a subgraph, such that the load is balanced equally among the leaders. Now, each of the controllers is responsible for 35 or 36 nodes, hence the number of workers assigned to each leader is much more balanced. The drawback is, that the path length and thus the latency on the communication channels between the leaders and their assigned workers can be higher. This can be seen in Figure 8(b) that shows the latency to nodes for the case of an assignment that optimises load imbalance.

Network planning. Another possibility for the optimisation of communication networks is network planning, i.e. changing the topology of a communication network. While in telecommunication networks, adding nodes can easily be accomplished by setting up hardware, in the context of enterprise social networks this would mean the addition of individuals (i.e. by hiring or transferring between groups). While this is not generally impossible, in the following we focus on targeted interventions by means of *adding links*, which can be achieved in a social organisation in much easier way, e.g. by asking people to collaborate or influencing enterprise social software in a way that it suggests contacts or communication. The addition of links in a social network can bring several improvements, however at the cost of additional expenditures. By adding a link between two persons, i.e., making them direct friends in the network, the path lengths between the existing friends of both persons might decrease leading to a shorter average path length in the network. Another benefit is that additional links can reduce the risk of a disconnection of a



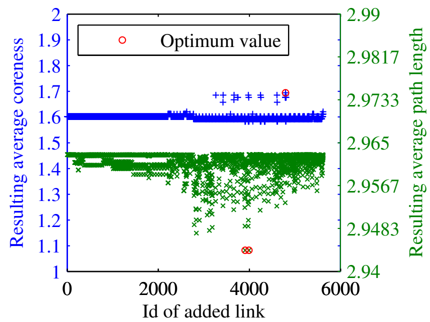
(a) Lowest maximum latency (b) Lowest load imbalance

(c) Assigned members (lowest max. latency) (d) Assigned members (lowest load imbalance)

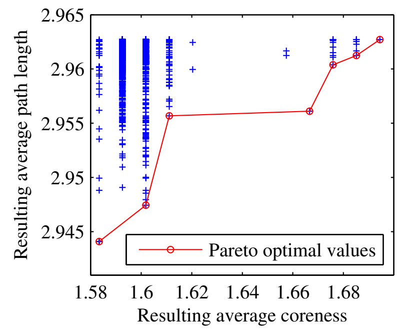
Fig. 8 Illustration of best controller placements according to different metrics.

Social Network Analysis in the Enterprise: Challenges and Opportunities 21

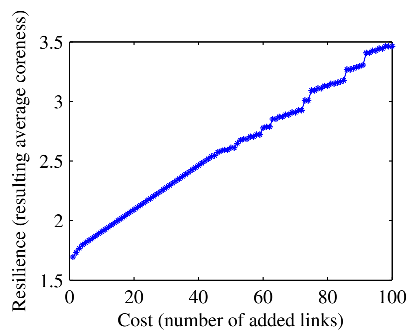
person or subnetwork from the rest of the network. A potential increase of the average coreness of nodes can be used as one possible measure to quantify the resulting increase in resilience.



(a) Exhaustive evaluation of all options to add one link.



(b) Trade-off between two objectives when adding one link.



(c) Stepwise addition of further links in a greedy manner.

Fig. 9 Visualisation of the improvement reached by adding additional links to the network.

Figures 9 exemplarily visualises the improvement reached by adding additional links to the GENTOO collaboration network. In Figure 9(a) all possible options of adding a single link to the network have been compared regarding the decrease in average path length (right y-axis) and the increase in average coreness (left y-axis). The optimum links according to both objectives have been highlighted. Two main findings can be observed: First, there is a high optimisation potential but the improvement reached by adding a link highly depends on the chosen link. Second, the best links to add regarding the optimisation of average path lengths (which can be seen as a proxy for communication efficiency) *do not* correspond to the best ones regarding the average coreness (which can be interpreted as a measure for resilience). In Figure 9(b) another visualisation is shown that substantiates the latter finding.

For all options to add a single link, on the x-axis the average coreness resulting from adding this link is shown, while on the y-axis the resulting average path length is shown. Each symbol in the plot corresponds to one or several links that lead to the same values on x- and y-axis. The highlighted symbols show the Pareto-optimal results regarding both metrics. Looking at Pareto-optimal results is a formal way to identify trade-offs in multi-criteria optimisation.

Finally, to look at the possible improvements when adding several links, a simple experiment has been conducted regarding the average coreness in the network. Subsequently, more and more links were added to the network. In each step of the iteration, all options to add a single link were tried and the option leading to the best improvement was chosen. Figure 9(c) shows for up to 100 added links the reached improvement in terms of average coreness. The linear relationship between resilience and cost highlights a trade-off to be decided by the network provider – or the company when looking at enterprise social networks.

4.3 Optimising Enterprise Social Networks: Research Challenges

Realising resilience in enterprise social networks and applying mechanisms from the design and planning of telecommunication networks poses a set of research challenges. These challenges arise from the scale of enterprise social networks, the applicability of the mechanisms, but also from the temporal dynamics of the network topology. In particular, it is not yet well researched how to model enterprise social networks and how to apply the well-known network communication methodology in the domain of enterprise social networks. In this section, several research questions are discussed, whether and how social network analysis may be beneficial to identify and react to problems in the enterprise in advance. Further, we will address the optimisation of an enterprise social network to improve resilience, effectiveness, and job satisfaction. However, beyond the technical aspects, the derived models and applied mechanisms also lead to ethical issues, which we discuss in a separate paragraph.

Modelling Social Systems. Social network structures in enterprises can be modeled in many ways and an appropriate representation has to be chosen depending on the relations and interactions within the company. For example, links can be created which resemble *boss-of*-relations (i.e. hierarchical structure) within a company or *communicates-with*-relations (i.e. actual communication structures) within the enterprise. Each network will have a specific structure depending on, e.g. enterprise hierarchy or communication flows, some of them are better suited for one company than for another. Thus, the question arises how to model the social network structure which is useful to identify problems or to optimise the network.

In every company there are persons which are more or less important. How should these personal attributes be modelled and how can their the workload and productivity be quantified? How can team work be integrated into the model? A company does not only need workers who are very targeted and finish one task after another. A company also needs workers that spend time socialising, finding new contacts, and connecting people. Such workers are essential for an efficient working atmosphere. How can their contribution be quantified? Which mixture of personalities is best for the company? Central nodes, for instance, are important to connect persons from different departments and to spread information, but might also suffer from high workload and stress. How can important nodes be identified? How can overloaded workers be detected and unburdened?

If an important person leaves the company, is ill for several weeks, or is moved to a different department, it can have bad consequences for the company like less effectiveness or productivity. What approach can detect and quantify such pending problems? If an enterprise has capacities to employ new workforces, the question is how new nodes are integrated, i.e. to which persons they are linked in the network? Are there other means to change the structure of the enterprise social network? Facilities (e.g. staff rooms) or events (e.g. company outings) support the dialogue of employees which are not directly working together. Thus, new links are created in the social network which foster serendipity and creativity especially when people of different disciplines get together. How can these means be modelled? How can the effect of such means be quantified?

Optimisation Capability. When optimising enterprise social networks with resilience mechanisms new challenges arise from the size of enterprise social networks, the temporal dynamics of the network topologies, but also from the means of modifying the network. Existing communication network heuristics for optimal solutions are typically limited to static networks, which - as has been shown in a recent line of research on temporal networks [9,12,13] - can differ significantly from actual communication flows that are due to the temporal ordering of interactions. Moreover, the difference in the structure between communication networks and enterprise social networks makes it difficult to apply common heuristics. Approaches such as routing optimisation, resilience analysis, or network planning are especially efficient in networks where the average node degree and connectedness of the network is already quite high. Furthermore, if personal interests and preferences of the employees are dominant and cannot or must not be influenced, the network structure is fixed. Other mechanisms, such as different message routing, can still improve the resilience of the social network. Instead of just processing plenty of requests, a new working directive could instruct central nodes to forward requests to different communities. Thus, load is taken of the central node and collaboration is enabled between the communities. Finally, compared to telecommunication networks, in social networks it is often more difficult to change or add links or nodes. Here again new challenges arise. Which real-world human resource actions resemble which communication network optimisation? What (side) effects occur when applying such actions in an enterprise? Does the gain of such actions exceed their costs? How do employees' personal attributes change when such actions are enforced?

Ethical Issues. Finally ethical issues have to be considered. It is important to understand that the analysis does not deal with abstract nodes but with humans.

Thus, it remains an open issue if a person can be judged by the structure of its social network projection. Moreover, in order to assess personal properties like work capacity or interaction capacity, working behaviour will have to be measured and captured in statistics. However, complex human personalities cannot be fully assessed by such statistics and a supervised working environment will induce changes in attitude and behaviour of workers.

If the social network graph is based on the communications within an online enterprise social network tool (e.g. email communication or Intranet platforms) where network structure and interactions can be easily obtained, different communication or collaboration channels, like face-to-face communication, cannot be captured. The question arises to which extend people are then discriminated who do not extensively use such a tool? Also [14] stated that "in contrast to conventional office software, micro-blogging implies social interaction and self-disclosure. This applies to social software in general. As a result, bringing applications like micro-blogging into the workspace goes beyond traditional technology acceptance theory."

When analysing an enterprise social network, workers might be identified whose working capacity is insufficient. Then, the optimisation of the network can lead to dismissal or demotion of workers. Moreover, forced human resource actions like moving persons to another department might have complex impacts on the employees' motivation. Also other resilience means might encounter employees' resistance. For example, the expected establishing of connections between different workers might not be accepted as it overrides the workers' own preferences for selecting social peers. Thus, it remains an open question which actions can - and should - be enforced by the company without running into ethical issues.

5 Conclusion

The resilience and efficiency of communication networks is a major topic in the network research communities both studying social and telecommunication networks. With the rise of collaboration platforms in enterprises social network structures on top of technical systems emerge which reflect the social organisation of an enterprise. Therefore it is tempting to utilise known results and insights from the optimisation of telecommunication networks. This maybe helpful for enterprises to improve their human resource management by pre-emptively taking load of busy and central workers and improve the social network structure to increase information diffusion and to accomplish a healthy work environment. The main contributions of this chapter are a) to summarise network-theoretic measures and interpret their meaning in the context of enterprise social networks, b) to illustrate how enterprise social networks can be monitored by showing an example from OSS communities, and c) to demonstrate approaches from the optimisation of telecommunication networks and to apply them to a real-world collaboration network.

Our results highlight new technical, scientific and ethical challenges which arise when wanting to monitor and optimise enterprise social networks. Combining expertise from the modeling and analysis of complex networks, the design and optimisation of telecommunication networks, as well as from the social sciences, the emerging interdisciplinary field of socio-informatics has the potential to address these challenges.

References

1. Cohen, R., Erez, K., ben Avraham, D., Havlin, S.: Breakdown of the internet under intentional attack. Phys. Rev. Lett. **86**, 3682–3685 (2001). DOI 10.1103/PhysRevLett.86.3682. URL http://link.aps.org/doi/10.1103/PhysRevLett.86.3682
2. Dunbar, R.: Grooming, gossip, and the evolution of language. Harvard University Press (1998)
3. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Mathematical Journal **23**(2), 298–305 (1973)
4. Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry **40**(1), pp. 35–41 (1977). URL http://www.jstor.org/stable/3033543
5. Garcia, D., Mavrodiev, P., Schweitzer, F.: Social resilience in online communities: The autopsy of friendster. In: Proceedings of the First ACM Conference on Online Social Networks, COSN '13, pp. 39–50. ACM, New York, NY, USA (2013). DOI 10.1145/2512938.2512946. URL http://doi.acm.org/10.1145/2512938.2512946
6. Garcia, D., Zanetti, M.S., Schweitzer, F.: The role of emotions in contributors activity: A case study of the gentoo community. In: Proceedings of the 3rd International Conference on Social Computing and Its Applications (SCA2013) (2013)
7. Hock, D., Hartmann, M., Gebert, S., Jarschel, M., Zinner, T., Tran-Gia, P.: Pareto-Optimal Resilient Controller Placement in SDN-based Core Networks. In: 25th International Teletraffic Congress (ITC). Shanghai, China (2013)
8. Hock, D., Hartmann, M., Menth, M., Schwartz, C.: Optimizing Unique Shortest Paths for Resilient Routing and Fast Reroute in IP-Based Networks. In: IEEE/IFIP Network Operations and Management Symposium (NOMS). Osaka, Japan (2010)
9. Holme, P., Saramki, J.: Temporal networks. Physics Reports **519**(3), 97 – 125 (2012). DOI http://dx.doi.org/10.1016/j.physrep.2012.03.001. URL http://www.sciencedirect.com/science/article/pii/S0370157312000841
10. Newman, M.E.: Networks: An Introduction. Oxford University Press (2010)
11. Olguin, D., Waber, B., Kim, T., Mohan, A., Ara, K., Pentland, A.: Sensible organizations: Technology and methodology for automatically measuring organizational behavior. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **39**(1), 43–55 (2009). DOI 10.1109/TSMCB.2008.2006638
12. Pfitzner, R., Scholtes, I., Garas, A., Tessone, C.J., Schweitzer, F.: Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. Phys. Rev. Lett. **110**, 198,701 (2013). DOI 10.1103/PhysRevLett.110.198701. URL http://link.aps.org/doi/10.1103/PhysRevLett.110.198701
13. Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Juan Tessone, C., Schweitzer, F.: Slow-Down vs. Speed-Up of Diffusion in Non-Markovian Temporal Networks. ArXiv e-prints (2013)
14. Schöndienst, V., Krasnova, H., Günther, O., Riehle, D.: Micro-blogging adoption in the enterprise: An empirical analysis. In: Wirtschaftsinformatik, p. 22 (2011)
15. Serrano Zanetti, M., Scholtes, I., Tessone, C.J., Schweitzer, F.: The rise and fall of a central contributor: Dynamics of social organization and performance in the gentoo community. In: Proceedings of the 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE 2013), held at ICSE 2013 (2013)
16. Van Mieghem, P.: Graph Spectra for Complex Networks. Cambridge University Press (2012)
17. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press (1994)

18. Wu, C.W.: Algebraic connectivity of directed graphs. Linear and Multilinear Algebra **53**(3), 203–223 (2005)
19. Zanetti, M.S., Sarigöl, E., Scholtes, I., Tessone, C.J., Schweitzer, F.: A Quantitative Study of Social Organisation in Open Source Software Communities. In: A.V. Jones (ed.) 2012 Imperial College Computing Student Workshop, *OpenAccess Series in Informatics (OASIcs)*, vol. 28, pp. 116–122. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2012). DOI http://dx.doi.org/10.4230/OASIcs.ICCSW.2012.116. URL http://drops.dagstuhl.de/opus/volltexte/2012/3774
20. Zanetti, M.S., Scholtes, I., Tessone, C.J., Schweitzer, F.: Categorizing bugs with social networks: a case study on four open source software communities. In: 35th International Conference on Software Engineering, ICSE '13, San Francisco, CA, USA, May 18-26, 2013, pp. 1032–1041 (2013)