RESEARCH PAPER

Identification of User Roles in Enterprise Social Networks: Method Development and Application

Janine Hacker · Kai Riemer

Abstract The importance of gaining insights into informal organizational structures for management purposes is acknowledged by both research and practice. However, "traditional" approaches to analyzing informal organizational social networks involve significant manual effort and do not scale for larger datasets. Enterprise Social Networks (ESN) have emerged as important tools for informal employee interactions, such as for problem-solving and information sharing. While the analysis of ESN back end data might provide insights into the informal fabric of organizations, and in particular employees' roles in such networks, there is a lack of systematic approaches for carrying out ESN analytics, such as for user role identification. Following a design science research process, a process-based method to identify user roles from ESN data was developed and evaluated. The method's efficacy is demonstrated through an in-depth application in a case study of Australian professional services firm Deloitte. In doing so the paper shows how ESN data can be utilized to derive metrics that characterize participation behavior, message content, and structural network positions of ESN users.

Keywords Social software · User roles · CRISP-DM · Role analysis · Knowledge management · Design science research methodology

J. Hacker (✉)
Institute of Information Systems, University of Liechtenstein, Fürst-Franz-Josef-Strasse, 9490 Vaduz, Liechtenstein
e-mail: janine.hacker@uni.li

J. Hacker
Institute of Information Systems, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lange Gasse 20, 90403 Nuremberg, Germany

K. Riemer
Discipline of Business Information Systems, The University of Sydney, Abercrombie Business School Building H70, Darlington, NSW 2008, Australia
e-mail: kai.riemer@sydney.edu.au

1 Introduction

Who do you typically turn to for help at work, to find a piece of information, or for exchanging ideas? And in which situations would colleagues ask you for advice rather than someone else? As employees draw on personal contacts to perform their day-to-day work, informal relationships emerge as the crucial "company behind the charts" (Krackhardt and Hanson 1993) that sits alongside the formal structure of an organization (Allen et al. 2007). One central theme in investigating such informal organizational networks is the identification of key roles, such as *central connectors* or *boundary spanners* (e.g., Cross and Prusak 2002). For managers, insights into the informal fabric of the organization can inform staffing decisions, retention management, talent management, as well as succession planning (e.g., Parise 2007). It is generally acknowledged that fostering insights into informal networks emerging in organizations is of relevance to both research and practice (e.g., Parise et al. 2006; Fischbach et al. 2009; Cross et al. 2013). At the same time,

"traditional" approaches to analyzing informal organizational structures involve significant manual effort (e.g., Steiny and Oinas-Kukkonen 2007).

Enterprise Social Networks (ESN) support the communicative and social aspects of work within organizations by providing a space for conversations (Riemer and Scifleet 2012). ESN use cases, among them the generation of new ideas, information sharing, and problem-solving, prove the value of ESN typically associated with informal structures. Since ESN user activities are available as digital traces in the form of messages and meta data stored in the system back end (Behrendt et al. 2014a), the analysis of such digital traces offers promising new ways for identifying informal communities, information and communication flows as well as influential users (Leonardi et al. 2013). Yet, while previous ESN research emphasizes the need to better understand user roles in the context of knowledge work (e.g., Leonardi et al. 2013; Hacker et al. 2017a, b), there are only very few studies performing empirical analyses of ESN user roles (e.g., van Osch et al. 2016a; Cetto et al. 2018).

One reason is the lack of a comprehensive method that takes into account the specifics of the ESN context and that would guide researchers (and practitioners) in performing ESN user role identification in a systematic way. We argue that this gap impedes the advancement of the emerging field of ESN analytics. Consequently, we investigate the following question with the view to deriving a suitable method: *How can ESN back end data be utilized and analyzed in a systematic way to identify user roles, emerging through interactions on ESN platforms?*

To address this question, we develop a method for carrying out a structured analysis of ESN data for identifying user roles. We term this method the ESN User Role Identification Process (EURIP). The development of EURIP follows the steps of the design science research methodology (DSRM) process model by Peffers et al. (2007). Content-wise, it draws on prior work identifying informal organizational structures (e.g., Cross and Prusak 2002; Allen et al. 2007), existing literature on user role identification in public online communities (Chan et al. 2010; Rowe et al. 2013), as well as the emerging body of dedicated ESN research (e.g., Oettl et al. 2018; Cetto et al. 2018). Process-wise, we utilize the cross-industry standard process for data mining (CRISP-DM) reference model (Chapman et al. 2000) as a generic template, which we instantiate and adapt with ESN analytics-specific activities to derive the EURIP.

We demonstrate application of EURIP and evaluate the model with a detailed case study of the Australian professional services firm Deloitte. We performed a cluster analysis on Deloitte's ESN data that identified nine distinct user roles capturing different work-related communication behaviors. Successful application of the method, identification of user roles, and feedback from the company on the usefulness of these roles, ensures positive evaluation of EURIP.

The contribution of our study to Information Systems (IS) literature is threefold: Firstly, we contribute to the field of informal organizational networks an alternative method to identify (more differentiated) organizational roles. Secondly, we contribute to the emerging field of ESN analytics (Riemer et al. 2018) EURIP as a systematic and comprehensive approach to identify ESN user roles, as well as a set of novel metrics. Thirdly, and as a by-product to research, we contribute to ESN research more generally a unique set of ESN user roles derived from a real-life case. In addition, we contribute to practice a method for identifying ESN user roles to inform decisions in areas such as management of knowledge work and human resources (HR), where the identification of critical knowledge resources remains an important managerial challenge.

2 Overview of the Design Science Study

We follow a design science research (DSR) process. Having become a cornerstone of IS research (Hevner et al. 2019), DSR aims to create, and theorize the creation of, new and innovative artifacts (Hevner et al. 2004), comprising IT systems, as well as constructs, models, methods, and instantiations of such artifacts in application environments (March and Smith 1995). In this research, we engage in the development of a new method. In doing so, we follow the research process articulated by Peffers et al. (2007), termed the DSRM process model, which embodies the principles and guidelines of DSR (e.g., Hevner et al. 2004). As such, it provides a structured and robust framework for carrying out, evaluating, and presenting DSR in IS research.

The DSRM includes six steps in a nominal sequence, each of which we follow in this study, namely (1) motivation and problem identification, (2) definition of a solution's objectives, (3) design and development of the artifact, (4) demonstration of its application, (5) evaluation of its efficacy, and (6) communication of results from the study. All steps of the DSRM have been conducted in close interaction with our case company Deloitte Australia (see Sect. 6.1): Having identified a lack of supportive methods to conduct ESN user role identification in the literature, members of the case organization were presented the idea of performing this kind of analysis and to develop the EURIP. Interviews and consultations with members of the case organization involved in knowledge management (KM), HR, and innovation management activities were conducted to receive feedback on and refine initial ideas as

Table 1 DSRM adopted from Peffers et al. (2007)

DSRM activity	Activity description	Application in this study
Problem identification and motivation	Identify and describe the importance of the problem	• Motivation of the study: Informal social networks are valuable for organizations, which will benefit from gaining a better understanding of actors assuming key roles in such emerging networks
		• ESN offer novel opportunities for analysis and identification of user roles
		• There is a lack of systematic approaches for carrying out such user role identification
Definition of objectives for the application	Define the artifact that accomplishes the objectives	We propose the development of EURIP, which aims to facilitate the conduct of ESN user role analysis. As detailed objectives we stipulate that EURIP should:
		• Produce interpretable results
		• Facilitate the discovery of a differentiated set of roles
		• Produce results that correspond with prior research
		• Produce roles that users can identify with
		• Produce results decision-makers find useful
		• Build on existing, best practice principles
		• Include a tailorable catalogue of metrics
		• Be comprehensive yet flexible to allow user role identification in various organizations
Design and development	Identify the requirement for the solution and create the artifact	• Design of EURIP is based on prior works in the fields of organizational Social Network Analysis (SNA), online social spaces, and ESN; and considers existing data analysis processes
Demonstration	Use artifact to solve problem	• Detailed application of EURIP in a suitable case study demonstrates applicability, and that it produces suitable results as judged against objectives
Evaluation	Iterate back to the design of the artifact	• Evaluation of EURIP against requirements and objectives, and identification of areas of improvement
Communication	Publish and communicate the results	• Results from the study are intended to be communicated via this paper

well as to come up with suitable objectives for the design and development of the EURIP from a practical point of view. The case organization provided ESN data for demonstrating the EURIP and supported the evaluation of the EURIP by participating in further interviews and workshops. Details on the case company's involvement can be found in the respective sections throughout this article.

The remainder of this paper is structured as follows: Motivation of the study and problem identification, presented in Sect. 3, encompasses an analysis of prior work and the identification of a gap in the literature, which allows articulating the design research problem. In Sect. 4, we translate this into concrete objectives for the EURIP, the method to be developed subsequently. Section 5 presents the design of the EURIP utilizing prior work. The EURIP is then applied in a case study, detailed results from which are reported in Sect. 6. Our evaluation of the method against the set of requirements identified earlier is presented in Sect. 7. Table 1 provides an overview of how the DSRM is applied in this study to motivate, present, demonstrate, and evaluate the EURIP.

3 Motivation and Problem Identification: Prior Work on Employee and User Roles

In this section we characterize the problem space targeted by our study. Using prior work on user role identification in different contexts, we establish the relevance of, and motivate, our method development, as well as derive a conceptual basis for its design. We argue that organizations will benefit from a structured method for the systematic identification of user roles from ESN data, and locate our method within the emerging field of ESN analytics, a subfield of social media analytics and related more generally to research employing organizational SNA to identification of network structures and user roles in organizations. We begin by establishing the relevance of informal networks in organizations, before we provide an overview of prior work in three areas: (1) traditional approaches to identify networks and roles in organizations, (2) user role identification in public internet communities, and (3) user role identification in the context of ESN itself. We end the section with a concrete problem statement motivating the method development, and positioning EURIP as a method for ESN analytics.

3.1 Motivation: The Value of Informal Organizational Communication Networks

Informal organizational networks are "the networks of relationships that employees form across functions or divisions to accomplish tasks fast" (Krackhardt and Hanson 1993, p. 104). They enable work and are crucial to the innovativeness of companies (Brown and Duguid 2001; Allen et al. 2007). Informal relationships between employees reflect personal relationships that are formed through friendship, acquaintanceship or between colleagues (Chan and Liebowitz 2006). Even though such relationships are typically formed in face-to-face settings, technologies supporting employee interactions allow for similar relationships to emerge (von Krogh 2012), especially as work is becoming more and more distributed and virtual (e.g., Hacker et al. 2019). As for ESN, the emerging relationships may be reflective of or complement relationships in other organizational contexts (Riemer et al. 2015; Klier et al. 2017).

Organizational actors assuming particular roles within such networks, e.g. the role of a *boundary spanner* or *broker*, contribute significantly to the efficiency and effectiveness of the network and, hence, create value for an organization. While traditionally informal networks and embedded roles remained largely invisible, due to a lack of suitable approaches to collect and analyses the required (large) datasets (Fischbach et al. 2010), the analysis of ESN back end data represents an interesting alternative (Leonardi and Vaast 2017). We argue that the identification of ESN user roles with a focus on their communication behaviors will be useful to organizations and their decision-makers. Most generally, an awareness of roles that employees perform during their day-to-day work deepens the understanding of informal work practices as well as of the actual roles of people and their contributions (Parise et al. 2006), thus helping organizations to "know what they know" (Newk-Fon Hey Tow et al. 2012). In addition, such awareness will (1) aid in managing risks associated with different roles (Jennex and Durcikova 2013), for example to take countermeasures if a valuable employee is at risk of leaving, (2) in rewarding positive behaviors within the network, thus helping to establish an effective communication culture, and (3) in informing decisions concerning the selection of personnel, among them decisions in talent management (Whelan 2011) or succession planning.

3.2 User Role Identification using Organizational Network Analysis

Traditionally, informal organizational relationships have been mapped using survey data. For instance, employees are asked to indicate who they usually turn to for advice when they experience work-related problems (Cross et al. 2002). The data is then analyzed using SNA metrics, such as degree centrality, betweenness centrality, and closeness centrality (Freeman 1978), to reveal those individuals in key roles (e.g., Cross et al. 2001b). In doing so, studies identified roles such as *central connector*, *boundary spanner*, *information broker* or *peripheral specialist* (Cross and Prusak 2002; Parise et al. 2006).

Advantages of manual data collection methods are that the context of the investigated organization is usually well-known and that analyses focus on a particular question, e.g. how information on a certain topic is spread in the organization. On the downside, the reliance on self-reported data and the objective of keeping those questionnaires simple for the participants comes with certain problems (e.g., Steiny and Oinas-Kukkonen 2007). For example, survey participants are usually asked to disclose their main contacts only, which leads to a potential over-emphasis on important contacts and thus, incomplete networks. Most importantly, such approaches do not scale and hence are not feasible to analyze larger datasets.

3.3 User Roles in Public Online Social Spaces

The identification of user roles is an important means for understanding activity in online forums and discussion spaces (Gleave et al. 2009). Typically, the roles people assume are identified through their "structural signature", including an individual's structural position in the online community as well as behavioral patterns (Gleave et al. 2009). A combination of social network metrics (see Sect. 3.2) and measures quantifying user (inter)actions, for example *number of initiated threads*, are used to characterize user behavior. The metrics thus indicate distinct behavioral dimensions. An overview of user roles found in different types of online communities is provided in Appendix 1 (available online via http://link.springer.com).

In comparison to organizational SNA, role analysis in online social spaces is based on larger datasets that can be collected automatically. Data reflect a complete set of actual (virtual) interactions between community members. While this removes problems of self-reported data, it makes the subsequent steps of the data analysis more challenging. For instance, the relevance of features and metrics for a certain community and the meaning of relationships and positions in a community are often not evident. Indeed, since user behavior emerges in accordance with the specific context of an observed online community, it is up to the community analyst to establish suitable metrics and features to conceptualize and measure user behavior (Rowe et al. 2013). This issue is aggravated by the fact that making sense of analytical results is usually done by a team of researchers rather than jointly with

selected members of the community. While results are not transferrable to organizational contexts, and methods are not immediately applicable, such approaches are useful for the adaptation of certain metrics that might characterize more generic aspects of communication in electronic social networking platforms.

3.4 User Roles in ESN

ESN are online platforms that are accessible only by a particular set of users belonging to one organization (Leonardi et al. 2013). ESN allow users to present information about themselves, to actively contribute content by posting updates to the platform's discussion thread as well as to respond to the content created by other users, through commenting, liking or sharing (Behrendt et al. 2014b). Users can share files and mention topics or other users in their posts. User interactions may occur on the platform's main message stream, in groups, or using private messages (Leonardi et al. 2013).

All communicative actions of ESN users are accumulated in the platform's back end. Often referred to as digital traces, "digitally stored, event-based, chronological records of activities of actors, which result in direct or indirect actor relations or content in different data formats" (Behrendt et al. 2014a, p. 4), such data can be utilized to both, construct an activity-based network graph to perform SNA and to characterize user roles via particular communication behaviors. While data on *activities* (usage data) indicates, *how*, *when* and *where* an ESN activity was performed, data regarding *content* (user-generated data) describes *what* was posted, that is the content type (e.g., status update, file) and the content itself (e.g., a question). Data on *relations* (structural data) implies *who* interacted with *whom*, that is, the sender and target of directed messages (Behrendt et al. 2014a, b).

Prior research has utilized SNA to analyze interactions on ESN in order to reveal informal networks (Behrendt et al. 2014a), or to identify *value-adding users* (Berger et al. 2014). Yet, only a few studies have identified dedicated ESN user roles, notably Cetto et al. (2018), Frank et al. (2017), and van Osch et al. (2016a). For a comprehensive overview see Appendix 1. Beyond academic research, companies such as SWOOP Analytics Pty Ltd[1] (SWOOP Analytics 2019) offer ESN role analysis as part of their analytics portfolio (Riemer et al. 2018). Generally, studies using digital trace data employ a greater variety of metrics than survey-based studies, which goes hand in hand with finding more differentiated user roles. In contrast to organizational SNA, such roles also reflect aspects of platform adoption, usage frequency, and intensity.

3.5 Problem Identification: Toward a Method for Systematic ESN User Role Identification

Even though both academia and industry show an interest in ESN user role identification, there is currently no structured approach or process to perform ESN user role identification at the organization level. This lack may entail problems in the performance of ESN user role analysis as well as limit the validity and comparability of findings. Approaches employed in industry, e.g. by SWOOP Analytics Pty Ltd, are used across a large number of companies and provide comparability, but identified roles are generic and pre-defined and not derived to represent a particular organizational ESN. Table 2 summarizes the main contributions and shortcomings of prior work informing the development of the EURIP. Note that these are shortcomings mainly in so far as they limit direct transferability to an ESN context.

So motivated, we set out to develop EURIP as a systematic method for ESN user role identification, and a contribution to the emerging field of ESN analytics, which Riemer et al. (2018, p. 3) define "as methods and practices for the identification and utilization of metrics and models for measuring different aspects of user activity in ESN, including user activity levels and user profiles, network activity levels, structural network characteristics, and network health indicators, in support of organizational goals and outcomes."

For doing so, the prior work as presented above can be exploited to derive suitable metrics that will utilize ESN data and form the basis of this method. Table 3 lists sample questions relevant for the analysis of user behavior and role discovery in ESN according to different categories of digital traces suggested by Behrendt et al. (2014a). The translation of these and other questions into suitable metrics promises to lead to the identification of users with varying degrees of topical focus, thread initiation ratio, and levels of participation.

4 Defining Objectives: Specification of Outcome- and Process-related Requirements

Objectives for the development of EURIP pertain to both intended outcomes of method application, as well as more generally the quality of the concepts and activities that make up the process by which such outcomes are produced. Outcome-related objectives are those that the method will later be evaluated against for efficacy and suitability. Results produced by the EURIP will be suitable if they satisfy the following criteria:

---

[1] https://www.swoopanalytics.com/.

Table 2 Summary of contributions and shortcomings of prior work

Field	Contribution to this study	Shortcomings from the viewpoint of this study
Identification of organizational roles based on (offline) surveys	• Provides context: application of SNA in organizational context • Theoretical foundation for identifying informal networks and roles in organizations • KM-related key roles	• Survey-based data collection implies that reports by respondents are subjective • Potential over-emphasis of key contacts in the network • Limited capacity to collect data for entire networks, lack of scalability
Identification of user roles in public online social spaces from (digital) activity traces	• Metrics and approaches to identify user roles in online communities using (digital) activity traces • Relevant roles in different types of online communities	• Lack of interaction between researchers and community members across stages of role analysis project may affect validity of the role analysis results
Identification of user roles in ESN from (digital) activity traces	• Initial set of metrics and insights regarding data analytics methods to identify ESN user roles using (digital) activity traces • Initial insights regarding ESN user roles	• Lack of interaction between researchers and members of the case organizations across stages of role analysis project may affect validity of the role analysis results • Consideration of a limited set of dimensions/metrics to identify roles

- The results of the data analysis, e.g. the outcome of a cluster analysis, are interpretable by the analyst, as results are otherwise unfeasible (Schendera 2010, p. 131).
- The identified set of roles is differentiated enough to capture a variety of user behaviors, because without differentiated roles we would be unable to get in-depth insights into how a community functions (e.g., Angeletou et al. 2011; Füller et al. 2014)

- The results correspond, at least in key aspects, to prior research as a way to establish their efficacy.

In terms of general objectives of data analysis projects (e.g., Chapman et al. 2000), results produced by the EURIP should fulfil the following criteria:

- Users of the investigated ESN are able to identify with the roles obtained, because they are best able to cross-check and validate the results.

Table 3 Questions for analyzing ESN user behavior

Aspect	Information about …	Exemplary questions for analyzing ESN user behavior
Who?	Individual users	• How long has a user been registered on the platform? • What information can be drawn from the user's profile page?
To/from whom?	Target/sender of an activity	• How many different users did a user send messages to? • How many users sent messages to an individual user? • To what extent did a user engage with users with different backgrounds?
How?	Type of (inter)action	• How many status updates did a user post? • How many threads did a user start? • How many replies did a user write? • How many replies did a user receive?
When?	Time of (inter)action	• To what degree did a user continuously contribute to the platform? • How often did a user contribute to the platform? • How often did a user consume content on the platform? • At what time of day did a user consume content on the platform??
Where?	Place of (inter)action	• How many messages did a user post on the platform's main stream (in groups/using private messages)? • How many messages did a user read on the platform's main stream (in groups/using private messages)?
What?	Object of (inter)action Content object Content itself	• How many files did a user share? • How many questions did a user ask? • How often did other users request information from a user? • To what extent did a user provide others with valuable content? • What type of content did a user view?

- Corporate stakeholders, such as managers, find the roles useful for sense-making as the application of the EURIP should lead to insights that are relevant for them.

To achieve this, we argue that the design of the EURIP should firstly *make use of established principles* and "best practices" in deriving a usable and robust method. To this end, the design should draw on existing data analysis processes or process models to ensure the comprehensiveness and flexibility of the sequence of steps included in EURIP. Secondly, in terms of the content of the method, its design should *incorporate insights from prior work* in studies analyzing roles in both offline and online settings (see Sect. 3), while mitigating the weaknesses and shortcomings of existing approaches. In doing so, it should include guidelines based on previous work on how metrics can be designed and adapted. Furthermore, the design of the method should *allow for contextualization to a specific case environment*. This includes accommodating the application of different data mining techniques and different types of role analysis depending on the investigated company and dataset, to meet the requirements of multiple case organizations. Finally, the method should take into account the *close interaction between the data analyst and the case organization*, such as domain experts, since user roles emerge differently depending on the context of an organization (see Sect. 3.3), and incorporation of a thorough analysis of the initial situation at any case organization, as well as a reflection of (preliminary) findings with domain experts throughout the project, in order to derive relevant results.

5 Design: Development of the ESN User Role Identification Process (EURIP)

We assert that the identification of user roles from ESN back end data can be considered a knowledge discovery project. Such a framing allows drawing on standardized knowledge discovery processes (KDPs) as the basis for method development. It has been argued that the goal of KDPs is to discover useful patterns in data that ultimately contribute to the generation of new knowledge in a domain (Cios et al. 2007). To this end, KDPs typically comprise the entire knowledge extraction process, from data collection over data mining to results visualization.

Relying on a (standardized) process model to perform ESN user role identification ensures that such projects follow a cohesive and deliberate structure. A role analysis process is also a helpful tool to communicate the need, benefits, and steps of an ESN user role identification project (Cios et al. 2007).

Academic research models include the KDP model by Fayyad et al. (1996), which consists of *selection*, *pre-processing*, *transformation*, *data mining*, and *interpretation/evaluation*, as well as the data mining model by Runkler (2015), which differentiates *preparation*, *pre-processing*, *analysis*, and *post-processing*. Models commonly used in industry include SEMMA (*Sample*, *Explore*, *Modify*, *Model*, *Assess*) by the SAS Institute (2012) and the CRISP-DM reference model (Chapman et al. 2000). The latter includes six phases, *business understanding, data understanding, data preparation, modelling, evaluation,* and *deployment,* that result in a cyclical process[2] (Chapman et al. 2000). The CRISP-DM reference model has been recognized as the leading industry model; it has previously been adapted to fit the needs of specific knowledge discovery projects, among them SNA (Asamoah and Sharda 2015), social collaboration analysis (Schubert and Schwade 2017), and evidence mining (Venter et al. 2007). Because it provides a comprehensive structure and accounts for continuous interaction between analysts and domain experts, we selected the CRISP-DM reference model as the basis for EURIP. CRISP-DM thus serves as a template or blueprint which we instantiate in the development of the EURIP.

Following CRISP-DM, we propose six main phases for EURIP; the activities in each are spelled out in Table 4. We renamed some of the phases to adapt the model to the problem domain. Figure 1 (based on Chapman et al. 2000, p. 12) illustrates actors typically in charge of activities in each stage, the data and software applications to be used, as well as reports as outputs of each phase. We attend to each phase and its activities in the following sub-sections.

5.1 Case Understanding

The aim of the first phase is to obtain a solid understanding of the organizational environment in which ESN user role identification is to take place. This is necessary as a foundation for interpreting the results of the data analysis later in the process. For example, the ESN data analyst might engage domain experts in a workshop to collect *background information* on the ESN, e.g., regarding how the platform was introduced and how it is being adopted and used across the organization, as well as the *objectives* for why the organization wants to identify ESN user roles, as this will influence and guide how the analysis will be performed.

In preparation for the actual analysis, resources such as *data*, *software* and *hardware*, as well as *personnel* are required. The *data* required to implement the metrics

---

[2] See Chapman et al. (2000) for details on the six phases of CRISP-DM.

Table 4 Phases and activities in the EURIP

| Name of original CRISP-DM phase	Name of phase in EURIP	Activities
Business understanding	Case understanding	• Elicitation of case background, and establishment of objectives for ESN user role identification • Assessment of resources and constraints • Assessment of risks and contingencies
Data understanding	Data understanding	• Collection of an initial ESN dataset • Exploration of ESN data • Verification of ESN data quality
Data preparation	Data preparation	• ESN dataset selection • ESN network graph generation • ESN metrics design and implementation
Modelling	ESN user role identification	• ESN data sampling • ESN user role identification
Evaluation	ESN user role interpretation and assessment	• Assessment of findings from a business perspective
Deployment	Utilization	• Creation of a utilization plan
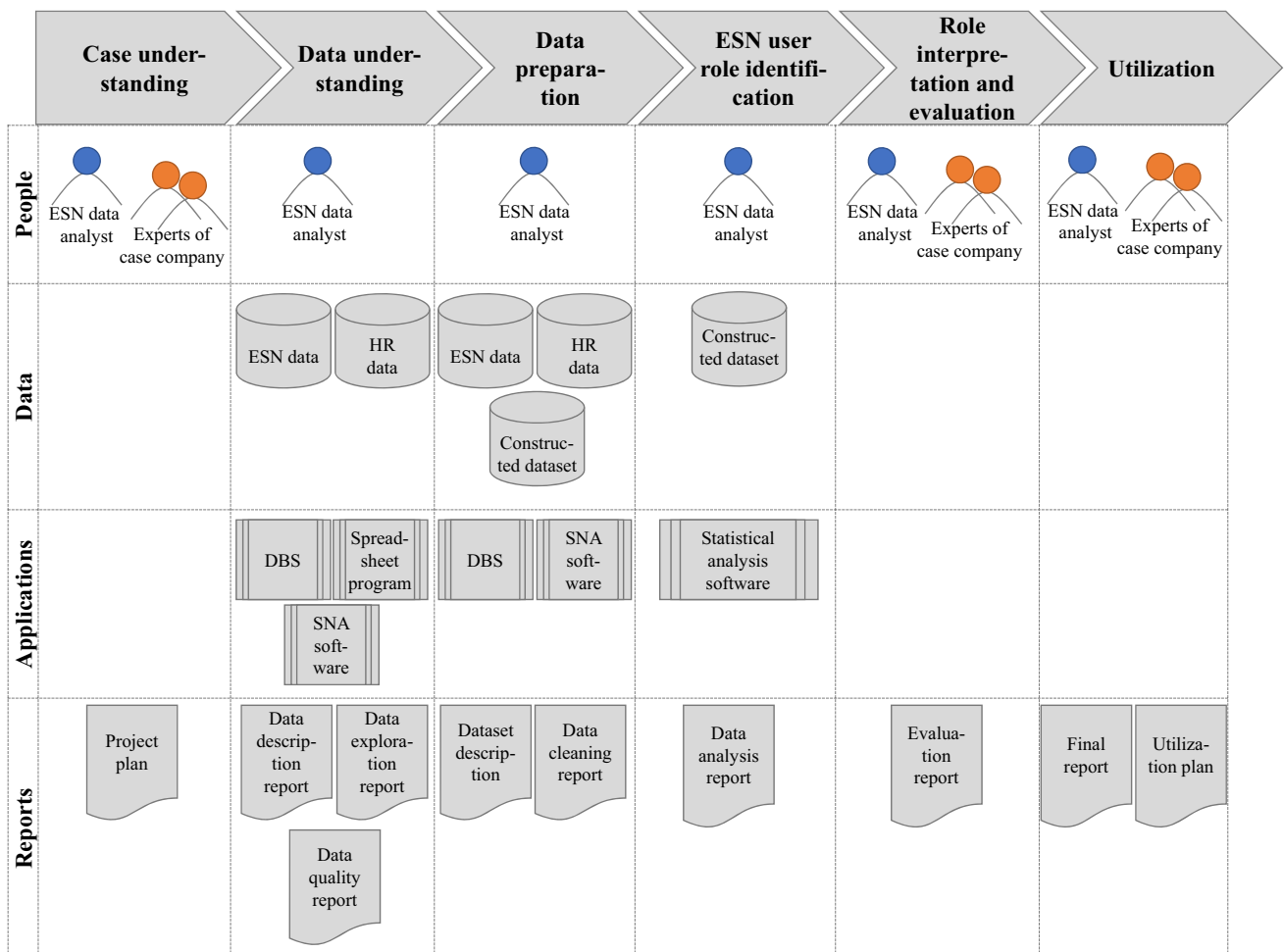
Fig. 1 Required resources and outputs of the EURIP (Based on Chapman et al. 2000, p. 12)

designed in step 3 (Table 3) may originate from different sources, such as the ESN back end (Behrendt et al. 2014a, b) or a company's HR system. In terms of employee-related (user-related) data, longitudinal data will be preferable to cross-sectional data as some attributes, among them a user position in the company hierarchy, may change over time. As for *computing resources*, ESN user role analysis requires software to administer and manipulate the data, such as a Database System (DBS) (e.g., MySQL) as well as software for statistical analysis (e.g., R), and SNA (e.g., Gephi). In terms of *personnel*, the ESN data analyst or the team of data analysts has to be familiar with a database query language, such as SQL, and SNA as well as have advanced statistics knowledge and experience in analyzing large datasets.

Finally, the ESN data analyst should identify risks that might affect the application of the EURIP (Chapman et al. 2000). Besides risks related to data access and quality, this should include an assessment of privacy-related and ethical concerns regarding the application of the method itself as well as using the results from the EURIP. For example, the ESN analyst and representatives of the case company should discuss who will get access to the results and especially whether and how individual employees would be presented with the results.

5.2 Data Understanding

In the second phase, the ESN data analyst will collect an initial ESN dataset from the organization and explore it in terms of the analysis objectives by calculating global metrics about messages (e.g., *distribution of messages over time*) and users (e.g., *distribution of users across hierarchical positions*). Finally, the analyst verifies the quality of the data by investigating its completeness, correctness, and missing values. For *completeness*, it is recommended to verify whether the user data contains both contributing and non-contributing users (e.g., those that have never actively posted on the ESN), as well as former users, or to check whether each message reply has a target message. To verify *correctness*, the analyst should track how data from

different sources, for instance, the ESN and a HR system, have been merged. Also, all tables should be checked for duplicates, implausible and missing values that might potentially compromise metrics calculation.

5.3 Data Preparation

The third phase comprises three sub steps: (1) selecting, formatting and cleaning the relevant ESN dataset, (2) constructing the social network graph from the ESN activity data, and (3) deciding on the range of metrics to be included in the analysis.

Having obtained an understanding of the particular dataset in the previous phase, the ESN data analyst will, depending on the analysis objectives, select a subset of the data spanning a certain period of time or focus on a subgroup of users. The selected data is then cleaned or converted in accordance with the results of the data quality check (see Sect. 5.2).

Of particular importance is the generation of the ESN network graph as the basis for calculating any social network metrics. The relevant question here is, what constitutes a relationship within the ESN? This is largely a choice driven by the analysis objectives. Different options exist for calculating network edges that are based on different assumptions, e.g. we might assume that relationships are constituted by users communicating with each other directly by replying to each other's messages, or already by being part of the same message thread (Table 5). Obviously, the resulting networks will vary significantly (Behrendt et al. 2014a). Hence, the ESN data analyst needs to carefully select applicable relationship(s) and extract those from the dataset.

Finally, this step includes compilation of the metrics catalogue for the analysis in step 4. This can involve a combination of new metric design for the purposes of a particular study or specific to the dataset, as well as the selection and adaptation of metrics from prior work. Generally, a comprehensive set of metrics covering a wide range of behavioral features should be compiled, in order to ensure variation in the data for surfacing suitable user

Table 5 Identifiable types of relationships based on ESN data (adapted from Hacker et al. 2016, p. 93)

Type of relationship	Description
Following	Edges between users are based on following another user's updates
Replies/comments	Edges between users are based on who comments on whose messages
Common conversation participation	Edges between users are based on the common participation in conversations (e.g. threads or discussions)
Users and topic tags	Edges between users and topics occur if users create or use a topic tag
User tags	Edges between users are based on who tags whom in a post
Questions	Edges between users indicate that a user asks another user a question

roles, e.g. based on the categories identified in Table 3. Metrics may reflect both contributing and reading activities, e.g. concerning when and where a user posts something or consumes existing content (van Osch et al. 2016a). Metrics used in previous studies on ESN user behavior (e.g., Holtzblatt et al. 2013; Cetto et al. 2018) (Appendix 1) might guide the development of the metrics catalogue. Building on our own prior work (e.g., Hacker et al. 2017a, b) we distinguish the following *types of metrics* that describe a range of activity suitable for identifying differentiated ESN user roles:

- *Activity metrics* characterize user behavior on the platform. They can be distinguished into *absolute metrics* and *average metrics*. Absolute metrics, such as *number of replies* created by a user, allow for counting how often a user engaged in a certain activity. Average metrics, such as *average number of replies created by a user per thread*, enable the comparison of information related to specific objects, such as threads, between users.
- *Social network metrics*, such as *in-degree* and *out-degree*, measure an individual's position within the ESN graph (Freeman 1978), which provide an indication of a user's influence and connectedness in the network.
- *Intrapersonal metrics* (Friemel 2008) are calculated from absolute activity metrics and put in comparison a user's specific behavior, for example initiating new conversations, with a user's overall activity on the ESN, thus measuring a user's preference for particular kinds of activity.

The ESN analyst will compile and select a set of metrics within these categories, which are then calculated for each user in the dataset and integrated into a format (e.g., a single table) as the basis for user role identification.

5.4 ESN User Role Identification

To decide on an appropriate user sample, the ESN data analyst will have to explore the constructed dataset by considering the distribution of basic variables, such as *number of messages*, across the user population. Considering the specifics of the provided dataset and the general level of participation, the analyst may specify a threshold for users to be included in the sample that facilitates meaningful analysis while keeping an adequate number of users. This is because for rather inactive users with only very few messages most metrics characterizing user activity will not be able to be calculated, or they will skew heavily.

Following the calculation of a range of metrics for each user, the identification of user roles will employ unsupervised learning methods, such as cluster analyses, to group users into categories (i.e., roles) based on similarities and differences in their behavior over time. The ESN data analysts will select one or more analytics methods, run several analyses, compare the quality of the obtained solutions, and select the most appropriate solution. Following user role identification, the distribution of sociodemographic attributes across the roles may be investigated to get a deeper understanding of who has been classified into particular role categories, and thus contribute to interpretation.

5.5 ESN User Role Interpretation and Assessment

The fifth phase involves the interpretation, assessment, and review of the findings together with decision-makers of the business functions involved in the research. The ESN data analyst verifies whether the discovered user roles are meaningful to the organization. If that is not the case or if particular user groups are not represented in the resulting roles, the steps of the data analysis have to be revised. Furthermore, the ESN analyst will discuss with participants how the insights can be applied in various organizational functions and at different levels. The interpretation and assessment of the ESN user roles as well as the application of the results can be performed by doing interviews with users of the ESN or key stakeholders who will be users of the results (e.g. managers), as well as in workshops with different representatives of the case organization.

5.6 Utilization

The final phase concerns the determination of a utilization strategy as well as the production of a final report. It will be specified whether role analysis will be performed again at a later point in time or at regular intervals. If the case company is interested in performing ESN data analysis continuously, the development of an application that executes the role analysis process (semi-) automatically is recommended. Furthermore, the utilization plan identifies relevant areas for applying the findings in the organization. The ESN data analyst may prepare the utilization plan based on insights from previous workshops and then refine and finalize it according to feedback collected in a dedicated workshop. The final report summarizes and organizes the steps and findings of the ESN user role identification project.

6 Demonstration: Application of the EURIP in a Corporate Case

In this section, we demonstrate and illustrate in detail the application of the EURIP with data obtained from an

Australian professional services firm. The research team was able to access a unique, large-scale dataset of the kind that would typically be available to an ESN data analyst in a corporate context. Successful application of EURIP will demonstrate that the method is capable of producing suitable, useful results.

6.1 Case Understanding

Our case study was carried out in cooperation with the Australian partnership of Deloitte Touche Tohmatsu. At the time of study, Deloitte Australia (hereafter Deloitte) had about 6000 employees located in 14 offices in Australia providing audit, economics, financial advisory, human capital, tax, and technology services. Deloitte was known as an early adopter of ESN. Often described as 'part of Deloitte's DNA', the particular ESN was a browser-based platform that offered a company-wide newsfeed, allowed users to create a profile, featured public and private groups, the sharing of messages and files, as well as communicating with others by commenting on their messages or the writing of private messages. Public groups could be viewed and joined by all network members whereas private groups were only visible to invited members. Deloitte started using the ESN in 2008.

Deloitte was interested in understanding better how individual employees share knowledge, communicate, and collaborate on the ESN. The study period was determined to comprise 1 July 2012 to 30 June 2013. At that time the ESN was fully adopted, had reached a state of maturity and was widely used in everyday activity. During the study period, the ESN was the main electronic communication channel besides an instant messaging tool for one-to-one exchanges, and more traditional media like telephone and email. As new employees had to create accounts as part of their on-boarding process, the set of registered users reflected well the composition of Deloitte's workforce. Due to the mature and wide use of ESN during the study period, the selected dataset presented an ideal case for demonstration of EURIP.

6.2 Data Understanding

Our ESN dataset comprised meta data of user messages with information on where a message was posted (in the main stream, a private or public group, or between two individuals), which message it replied to, if it was part of a thread (a conversation), whether it had an attachment, as well as the meta information mentioned below. All data was imported to a MySQL database.

We note that our dataset did not include data on consuming behavior (e.g. reading) because this ESN system did not collect such data. Also, for privacy reasons,

Table 6 Overview of the obtained ESN data

Aspect	Obtained ESN data
Who?	User ID (anonymized)
	Business unit; job title; location
To/from whom?	User ID (anonymized), message ID
	Business unit; job title; location
How?	Post a status update; create a thread; reply to a status update; receive a reply; tag another user; being tagged by another user
When?	Timestamp
Where?	Main stream; group ID; group type (public, private)
What?	Message ID
	Message type
	Keywords (praise, thanks, question words); tagged topics; attachments; length

Deloitte did not provide any actual message content or file attachments. To be able to develop metrics indicating the length of posts (e.g., Morzy 2009; Junquero-Trabado and Dominguez-Sal 2012), tagging of other users and topics (Richter and Riemer 2013), as well as behaviors related to question asking (e.g., Hansen et al. 2010; Burns and Kotval 2013), thanking (e.g., Graham and Wright 2014) and praising others (Richter and Riemer 2013), the company extracted additional meta information from message contents via certain keywords (see Table 6).[3] Additional information was obtained for a subset of users from Deloitte's HR system, such as gender, business unit, job title, and geographic location. This information would allow identifying the degree of homogeneity within a user's network of communication partners on the ESN (e.g., van Osch et al. 2016b). Table 6 provides an overview of the final dataset, structured along the conceptualization of digital traces by Behrendt et al. (2014a).

6.3 Data Preparation

Data cleansing and formatting included the deletion of automatically generated messages, the conversion of certain values, among them timestamps, the removal of duplicates, and the merging of sociodemographic information from the HR system into broader categories. The cleaned dataset comprised 61,945 messages posted by 3158 users, from a total of 6235 registered users. Figure 2 provides an overview of the kind of messages in the data set.

---

[3] As information regarding "likes" or points awarded for posts was not part of the data export, keywords pointing to one user praising another user (e.g., "well done") are also assumed to indicate the quality (e.g., Rowe et al. 2013; Füller et al. 2014) of a user's contribution or performance in general.
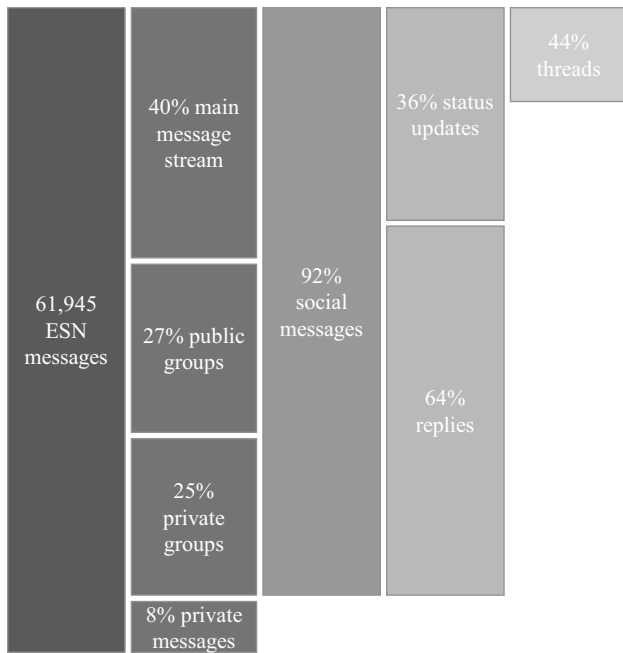
Fig. 2 Breakdown of ESN messages

To generate the ESN network graph, we decided to use direct reply-relationships (who replied to whom) as this type of relationship provided the most comprehensive

picture of user interactions. We then selected and calculated a range of metrics in the three categories *activity*, *social network* and *intrapersonal* metrics. Social network metrics were calculated with the statistical tool R (R Core Team 2015) and its *igraph* package (Csardi and Nepusz 2006). As the intrapersonal metrics were specifically generated for this case data, we present the list in Fig. 3. In total, we developed an initial set of 74 absolute metrics, 13 average metrics, 35 intrapersonal metrics as well as five social network metrics. Metrics were developed in an iterative manner, meaning that we added new metrics to the existing set of metrics as we discovered interesting aspects worthy of deeper investigation while exploring the data.

We then consolidated the number of metrics to a total of 47 metrics (see Appendix 2) by eliminating those that reflected redundant information when they could be calculated from each other. This was important to ensure that certain aspects of user behavior do not become over-represented (Hair et al. 2014, p. 456) in the subsequent cluster analysis (Sect. 6.4). Because they take into account different kinds of ESN digital traces (Behrendt et al. 2014a), the 47 metrics provide comprehensive coverage of different aspects of ESN user behavior, as well as a user's structural embeddedness and the diversity of users'

Fig. 3 Overview of the 35 intrapersonal metrics (Note that the metrics in the boxes shaded in light grey represent the numerators of the derived ratios, while the metrics in the boxes shaded in dark grey represent the respective denominators, such as *#threads created/#initial messages created (*included as *%createdThreads* in Table 7 respectively).)

communication partners in terms of business unit, hierarchy, and location.

6.4 ESN User Role Identification

Data analysis involved two steps, an initial principal component analysis (PCA) to further reduce the number of variables to a smaller set of linearly uncorrelated composite variables, and the cluster analysis. We note that PCA is recommended when dealing with potentially correlated variables (Schendera 2010, p. 19). We then performed a cluster analysis based on Ward's method, further optimized using K-means clustering. Since PCA does not recognize measurement error, the components obtained from PCA were not interpreted but merely used to replace the initial variables for the cluster analysis (Schendera 2010, p. 192). The interpretation of the clustering results was then done with the original variables respectively. We note that this analysis configuration, while in our view typical and fit for purpose, mainly serves to illustrate how user role identification can be done with EURIP, while other techniques, such as rule-based approaches (e.g., Rowe et al. 2013), can also be used.

6.4.1 ESN Data Selection and Sampling

This step involved (1) the selection of a user sample, (2) selection of metrics for conducting PCA and cluster analysis, (3) an outlier analysis, and (4) performing the PCA.

For inclusion in our *sample*, users were required to have sent enough messages to exhibit at least some behavioral patterns. We thus only considered users who had written at least twelve social messages during the study period, i.e. on average one message per month. Due to missing values, we excluded 10 users to arrive at a total of 899 users.

Next, *variables* (metrics) had to meet certain formal requirements. Beyond being relevant and reflecting unique information (see Sect. 6.3), the values calculated for a certain metric had to show a sufficient variance across the user sample (Schendera 2010, p. 293). We considered variance too low if more than 33% of the users had the same value. 11 of the 47 metrics were thus excluded. Moreover, metrics had to be correlated (Backhaus et al. 2016, pp. 298). Since the metrics were not normally distributed, we created a correlation matrix using Spearman's rank correlation coefficient (Sachs 1978, pp. 308) which was used to calculate Kaiser's MSA value for each variable and the correlation matrix. We found the MSA value of two variables to be below 0.5, which means that they were not suitable for extracting components (Backhaus et al. 2016, p. 398). After excluding this variable, a Kaiser's MSA value of ~ 0.71 for the correlation matrix indicated the correlation matrix to be suitable for PCA.

Next, we performed *PCA* using the remaining 34 variables. The optimal number of components was determined with a parallel analysis procedure which is implemented in the *nFactors* package (Raiche and Magis 2015). Nine components explaining 58% of the variance of the dataset were extracted.

Finally, we performed an *outlier analysis* on the 899 users in our sample to identify users with combinations of metrics deviating widely from other users. We created two ranking lists using single linkage (Schendera 2010, p. 25) and K-means clustering and excluded those users who belonged to the five percent that were assigned at last to a cluster according to both procedures. As a result of this analysis, 20 users were identified as outliers and removed from the sample, since outliers do not contribute to identifying typical clusters, though they might deserve a closer look by management individually.

6.4.2 ESN User Role Identification Using Cluster Analysis

With observations exceeding 250, non-hierarchical cluster analysis based on K-means was deemed suitable (Schendera 2010). However, a common problem of non-hierarchical methods is that the results may depend on the starting solution. This problem can be mitigated by using the results of a hierarchical clustering procedure, for example generated based on Ward's method, to specify the starting solution (Bortz and Schuster 2010, p. 461). In so doing, the advantages of both hierarchical methods and non-hierarchical clustering procedures can be usefully combined (Bortz and Schuster 2010, p. 461).

Moreover, the number of resulting clusters needs to be specified; we reasoned that this number should be between four and ten. A minimum of four ensures minimum differentiation of user roles, while a number greater than ten would challenge cluster interpretation. We performed several clustering iterations and determined the final number using the *NbClust* package (Charrad et al. 2014) which provides 30 indices to determine the relevant number of clusters. As a result, the number of clusters was determined as nine. Cluster analysis was then performed using the R functions *hclust* and *kmeans*, using the sample of 879 users and the nine principal components.

Having calculated the nine clusters, we evaluated the quality of the results. The average silhouette width of our clustering solution was 0.22 which is above the critical value of 0.2 and suggested a weak cluster structure (Martinez and Martinez 2005, p. 148). A test for cluster stability using a random starting solution (Schendera 2010, p. 132) revealed satisfactory stability. Goodness criteria aside, we note that the most important criterion is that the resulting clusters can be interpreted and named in a meaningful way,

and that they make sense in the context of the project (Schendera 2010, p. 131).

6.5 ESN User Role Interpretation and Assessment

To be able to interpret the clusters, we calculated the means of all metrics for all users belonging to each cluster (Schendera 2010, p. 131) (Appendix 3). We then checked for statistically significant differences in the variables between the different clusters using a Kruskal–Wallis test (Bühner and Ziegler 2009, p. 278). The last column of the table displayed in Appendix 3 displays the results of this test. Except for the variable *%UniqueTagsReceived*, all variables show highly significant differences at the p < 0.01 level between at least two clusters.

Based on the variables that show significant differences, we were able to interpret and name each cluster. To be considered as a defining characteristic for a cluster, the mean of a variable should represent an extreme value when compared to all other means. Table 7 lists selected metrics that are relevant for characterizing the nine clusters; green and red shadings have the highest significance for a particular cluster. In the following, we provide a brief description of each cluster using this information.

6.5.1 ESN User Roles

Cluster 1, Power users 4.7% of users, avg 288 messages. Due to their extraordinarily high number of messages and high percentage of replies, these users are highly connected. While they tend to focus their activities on the main message stream, they also contribute to a high number of groups and tag messages with various topic tags. Further, these users have the shortest delay in replying to messages and questions. In turn, they receive answers to questions quickly and attract above average replies. Given their high level of activity we term this cluster "power users". We expect these users to enjoy higher than average visibility and popularity in the network.

Cluster 2, Conversation starters 20.7% of users, avg 52 messages. Only 36% of these messages were posted in groups, which was the lowest proportion across clusters. These users further rank high in tagging other individuals and interacting with users belonging to different business units and different locations. Even though only 24% of their messages are initial messages, a high proportion (60%) attracts answers. Users in this cluster also post relatively many questions and participate in conversations that start with a question. We term this cluster "conversation starters", as these users seem to initiate many conversation threads of broad network appeal.

Cluster 3, Well-connected helpers 13.7% of users, avg 40 messages. These users also post messages predominantly to the main stream, and are well connected to users in other business units and positions. While these users post an average number of replies, messages are very long when in reply to questions. They tag other users most often and rank second in adding topic tags. As they are well connected to many different users, this indicates that they connect users with special requests to other users. These users also receive thank-you messages most often, which indicates they make valuable contributions. We name them "well-connected helpers" accordingly.

Cluster 4, Focused information sharers 9% of users, avg 41 messages. These users had the highest proportion of group messages but a low level of interaction with users in other hierarchical levels or offices. They tend to write initial messages as well as the first and last replies in a thread. They also include attachments and tag users. In turn, they are often tagged in initial messages and receive praise. While 69% of their messages are initial messages, only 31% of these yield answers. Given their high number of unsolicited messages (with attachments) in groups we name them "focused information sharers".

Cluster 5, Sporadic users 5.3% of users, avg 19 messages. These users show the lowest activity level and are named "sporadic users". They do not create many replies, and if so, with a very long delay. Interestingly these users are often tagged in questions, thanks or praise messages, even though they only receive few (directed) replies that are thanks or praise messages. This suggests that they receive thanks or praise for activity outside of the ESN.

Cluster 6, Task coordinators 21% of users, avg 52 messages. These users mostly post in groups (79% of messages). They are not well connected beyond their own business area, but interact well with users in different hierarchical positions. They post the longest questions to which they receive swift replies. They also rank first in tagging unique users in their messages and show the highest percentage of thank-you messages. Hence, we name the cluster "task coordinators" as users appear to engage in project work in their local areas.

Cluster 7, Offline experts 3.6% of users, avg 21 messages. Users in this cluster have the lowest scores for activities such as writing messages, creating threads and for tagging people or posting attachments. Only 17% of their messages are initial ones, with the majority being replies. Interestingly, the number of tags received exceeds the number of received replies and users are often tagged in questions which indicates a certain reputation outside of the ESN. They also receive many thank-you messages. Hence, we name them "offline experts" because they appear to only engage in the ESN when their expertise is requested, often by those in other offices and in other positions.

Table 7 Means of selected metrics per cluster (see Appendix 3 for all 47 metrics; green = highest mean, red = lowest mean across all clusters)

| | Cluster | | | | | | | | | |
	1	2	3	4	5	6	7	8	9	$\bar{x}$
N(879)	41	182	120	79	47	185	32	76	117	
Avg. Silwidth	0.45	0.26	0.14	0.14	0.18	0.29	0.26	0.10	0.20	
Who?/To and from whom?										
ClusteringCoefficient	0.03	0.08	0.11	0.10	0.11	0.08	0.14	0.07	0.11	0.09
%RepliesSentTo UserOtherBU	0.36	0.48	0.47	0.21	0.27	0.17	0.39	0.13	0.22	0.31
%RepliesSentTo UserOtherJT	0.63	0.65	0.64	0.42	0.51	0.65	0.75	0.13	0.66	0.58
%RepliesSentTo UserOtherLo	0.39	0.50	0.45	0.21	0.30	0.35	0.53	0.10	0.35	0.37
How?										
SocialMessages Created	287.51	52.47	40.19	40.81	19.36	52.11	20.63	35.61	24.52	52.53
%CreatedInitials	0.33	0.24	0.42	0.69	0.51	0.36	0.17	0.39	0.30	0.37
%CreatedThreads	0.46	0.60	0.41	0.31	0.48	0.56	0.01	0.56	0.58	0.50
AvgRepliesCreatedperThread	1.39	1.41	1.11	1.11	1.42	1.22	1.38	1.42	1.17	1.27
AvgRepliesReceivedperThread	4.47	5.70	3.22	2.12	4.43	3.68	0.03	4.14	2.91	3.78
%TagsSent	0.22	0.19	0.38	0.31	0.18	0.28	0.07	0.23	0.26	0.25
%TagsReceivedVsReplies	0.20	0.26	0.71	0.54	0.26	0.40	1.68	0.43	1.01	0.54
When?										
AvgTime_SendFirstReply (days)	0.63	1.33	1.06	1.94	349.48	1.24	34.60	6.03	5.34	22.06
AvgTime_ReceiveFirstReply (days)	1.01	0.77	0.79	2.11	1.36	1.03	353.59	0.98	1.90	14.00
TempConcentration	0.06	0.11	0.10	0.12	0.15	0.10	0.15	0.19	0.12	0.12
Where?										
GroupsContributedTo	12.76	4.18	3.66	2.94	2.26	3.78	2.69	2.88	3.00	3.89
%MessagesCreatedGroups	0.45	0.36	0.38	0.85	0.55	0.79	0.65	0.74	0.77	0.61
%PrivateGroupsContributedTo	0.22	0.21	0.25	0.45	0.19	0.54	0.31	0.54	0.50	0.37
What?										
%CreatedQuestions	0.20	0.42	0.17	0.15	0.35	0.31	0.09	0.32	0.06	0.26
%SentThanks	0.07	0.12	0.09	0.08	0.15	0.16	0.07	0.13	0.13	0.12
%SentPraise	0.03	0.02	0.04	0.03	0.01	0.04	0.04	0.02	0.08	0.04
%ReceivedThanks	0.08	0.08	0.13	0.13	0.05	0.10	0.13	0.05	0.11	0.10
%ReceivedPraise	0.02	0.02	0.05	0.07	0.03	0.04	0.01	0.03	0.06	0.04
%TopicTagsSent	0.05	0.06	0.11	0.07	0.05	0.05	0.12	0.04	0.05	0.06
%AttachmentsSent	0.18	0.11	0.22	0.38	0.22	0.16	0.08	0.12	0.13	0.17
%RepliesSentToInitialQuestion	0.39	0.51	0.39	0.24	0.46	0.43	0.42	0.45	0.28	0.41
AvgWordsperMessage	27.71	32.80	34.28	43.90	35.11	36.75	34.50	35.88	29.75	34.64

Cluster 8, Chat users 8.6% of users, avg 36 messages. Activities of these users show the highest temporal concentration, suggesting that they use the ESN quite selectively. Their initial messages are answered quickly. Of all clusters, they interact the least with users in other positions, business units, and offices. Hence, they mostly talk to their peers, underpinned by the fact that they send relatively many private messages, using the ESN more akin to an instant messaging tool. We thus term them "chat users".

Cluster 9, Team members 13.3% of users, avg 25 messages. While only 30% are initial messages, a high proportion of 58% attracts answers. Most messages (77%) are in groups, 50% of which are in private groups, but users in this cluster do not engage much in question and answer conversations. Also, their messages are the shortest of all clusters. On the other hand, they are often tagged in praise messages and praise others. Based on these characteristics, these users appear as typical followers, or "team members" engaged in their local divisional areas.

6.5.2 Assessing the ESN User Roles

In order to assess the identified user roles, we first compared them with prior work. We then carried out interviews with ESN users at Deloitte, and a workshop with decision-makers, to collect first-hand impressions regarding meaningfulness and usefulness.

Using existing literature, we checked for every identified user role if similar roles had previously been identified (see Sects. 3.2–3.4). As Table 8 shows six of the nine user roles are reflected in roles identified in other studies, while three of them, *task coordinators*, *chat users,* and *team members* did not show up in previous studies, and might be roles specific to either the ESN setting or the case company. We conclude that, since the discovered roles show similarity with previously identified roles, this lends support to the outcome produced by our method. We reiterate that the aim of our method is to enable organizations to identify specific roles that characterize their own user population. We would thus expect some roles to be quite unique to the case setting, while most will be typical in nature.

We further assessed the nine user roles in (14) one-on-one interviews with regular users of Deloitte's ESN to find out if the discovered behavioral patterns are relevant and perceived as distinct from the viewpoint of actual users. Specifically, users were asked to report on their own use and their observations of that of other users. They were then presented with the roles we identified and asked whether they could recognize these roles in the ESN community and assign themselves to one or more of the roles. Additionally, experts in charge of Deloitte's ESN were presented with the roles in a workshop and asked for

feedback. In general, interviewees and workshop participants were able to associate the roles with their real-life experiences using the ESN, which lends support to our developed metrics and interpretation of the clusters. Not surprisingly, most interviewees assigned themselves to not just one, but two or three of the roles that they enact to different extents or in different contexts, e.g. in the main stream as compared to groups on the ESN. We note that identified roles are ideal-types that result from the way *hard* clustering algorithms (as opposed to *soft* or *fuzzy* clustering methods) create non-overlapping clusters (e.g., Ferraro and Giordani 2015), whereby every user in the sample is assigned to one cluster only, even though they do show characteristics of other clusters as well. As such, the identified clusters represent average behavior profiles across the population with few users being actual ideal candidates that would naturally fall into one cluster only.

6.6 Utilization

We conducted two workshops with Deloitte representatives to discuss the user roles and potentials for their utilization. The feedback was that roles could support management of the ESN community by facilitating self-reflection and benchmarking of individual users, mentoring and coaching between users in different roles, as well as the monitoring of progression of users in roles perceived as vital over time. Especially the combination of results from ESN user role identification with employees' sociodemographic data, such as their tenure, position, business units, and location was suggested to facilitate additional insights. For instance, the company might compare the distribution of roles across different business units to find out about the extent to which employees share information and knowledge or collaborate.

In terms of strategic communications, knowledge about user roles was considered to be helpful to determine influential users, who, for instance, may usefully be employed as catalysts in certain corporate change initiatives. Especially users assigned to the clusters *power user* or *conversation starter* were deemed relevant. Also, participants discussed about the "right" role composition needed to keep up ESN engagement and maintain community health. Having too many users in roles that are mainly active in groups (e.g. *focused information sharer*, *coordinator*) might indicate the existence of knowledge silos and a lack of connectedness across the firm.

7 Evaluation of the EURIP

The last step in the DSRM (Peffers et al. 2007) is the evaluation of the design artifact against the objectives set

Table 8 Comparison of identified roles with roles from prior work

Case ESN user role	Corresponding roles from prior literature
(1) Power user	Active contributors (Holtzblatt et al. 2013)
	Answer persons (Viegas and Smith 2004)
	Central connectors (Cross and Prusak 2002)
	Super-promoters (van Osch et al. 2016a)
(2) Conversation starter	Discussion starters (Hansen et al. 2010)
	(Popular) initiators (Angeletou et al. 2011; Hacker et al. 2017b)
	Popular participants (Angeletou et al. 2011),
	Debaters (Hacker et al. 2017a, b)
	Contributors (Schwade and Schubert 2019)
(3) Well-connected helper	Boundary spanners (Cross and Prusak 2002)
	Experts (Welser et al. 2009)
(4) Focused information sharer	Focused (expert) initiator (Rowe et al. 2013)
(5) Sporadic user	Occasional users (Holtzblatt et al. 2013)
	Peripheral users (van Osch et al. 2016a)
(7) Offline expert	Peripheral players (Cross and Prusak 2002)
(6) Task coordinator, (8) Chat user, (9) Team member	No corresponding roles found

earlier in the process (see Sect. 4). For doing so, we carried out the Deloitte case study to demonstrate method applicability, as a form of observational evaluation according to Hevner et al. (2004). According to Venable et al. (2012), a case study is a form of ex post evaluation done in a naturalistic setting, which has several advantages, such as that it deals with real users and problems, which affords the best evaluation of effectiveness. In Sect. 4 we formulated a set of output-related objectives against which the developed method is to be judged:

- *Results have to be interpretable by the analyst*: We demonstrated that the method leads to interpretable results. We presented nine well-described ESN user roles, characterized in detail by particular ESN metrics.
- *The identified set of roles has to be differentiated to capture a variety of user behaviors:* The nine identified roles have clearly differentiated profiles, as indicated by how they are underpinned by certain metrics (see Sect. 6.5).
- *The results should correspond to prior research*: We have shown above that the core set of the discovered roles can be related to roles found in prior work.
- *Users of the investigated ESN have to identify with the roles obtained:* As discussed in Sect. 6.5, users were able to recognize themselves and others in the identified roles.
- *Corporate stakeholders, such as managers, should find the roles useful*: The interviews and workshop with different members of Deloitte revealed the findings obtained through the EURIP as useful for reflection and sense-making.

We conclude that the evaluation of EURIP through application in a naturalistic case setting has been successful since the method produced useful results from real ESN data, utilizing standard analysis techniques, as well as metrics derived from well-established categories to fit the unique characteristics of the data in the case.

In Sect. 4, we also outlined further process requirements by which the method was to be designed and the design elements that went into the method itself:

- EURIP draws on *established best practice principles* as it is built over the CRISP-DM reference model, which ensures a clear and logical structure.
- EURIP was designed by *incorporating insights from prior work* performing role analysis, in particular the catalogue of metrics used (see Appendix 2). While the actual metrics used in a case application have to be applicable to and calculated based on the particular ESN data set, the metrics catalogue will serve as a basis to do so.
- EURIP was designed to allow for *contextualization to a specific case environment*. As such, it is open towards using data analysis methods other than cluster analysis, such as contextualizing predefined sets of roles as presets for ESN user analysis (Schwade and Schubert 2019).
- EURIP incorporates *close interaction between analysis and organization* by suggesting various meetings of the ESN data analyst and members of the organization, particularly at the beginning and towards the end of a role analysis project. The activities in our case

application are merely indicative of the kinds of activities that might be usefully undertaken.

8 Conclusion

We developed a method for eliciting user roles from ESN data, which we named EURIP. We instantiated the CRISP-DM reference model as a template to develop a process-based approach that covers the entire KDP, facilitates continuous interaction between the ESN data analyst and members of a case company, and can be adapted to fit different objectives by amending the set of metrics, as well as analysis techniques, in a concrete application. We have demonstrated its efficacy and utility through an application in a case study of Australian professional services firm Deloitte. We make three distinct contributions to the literature, (1) to the field of informal organizational networks, (2) the emerging field of ESN analytics (Riemer et al. 2018), and (3) research on ESN more generally.

8.1 Contributions

Firstly, our study contributes to the literature on *informal organizational networks*. While this research highlights the usefulness of network-based approaches to the discovery of employees' roles in the informal structures of an organization (e.g., Cross et al. 2001a; Borgatti and Cross 2003), we pointed out certain limitations, such as the subjectivity of survey-based data collection (Steiny and Oinas-Kukkonen 2007) or the high manual effort involved (Fischbach et al. 2009). We contribute to this field a new/alternative method to identify organizational roles. The comprehensiveness and richness of ESN data facilitates the discovery of more differentiated and detailed sets of roles.

Secondly, to the emerging field of *ESN analytics* we add a new process-based method for identifying ESN user roles from ESN back end data, which provide a persistent record of user interactions over time. Leonardi et al. (2013) argue that ESN analytics can support the identification of informal relationships and communities as well as central actors in those emergent structures. Such insights are helpful to make employees and managers aware of "who knows what" and "who knows whom", which again assists in fostering collaboration and surfacing critical knowledge resources. While such insights are generally enabled by the transparency created through ESN itself, and can hence be gained by users who participate in the ESN, such transparency will necessarily remain partial and locally restricted to a user's field of vision, in particular in large corporate ESN. There is hence a need for systematic and comprehensive approaches to gain thorough insights that can inform managerial decision-making. With the EURIP we address this gap and encourage other researchers to conduct ESN user role identification and advance this stream of research. In addition, we provide as a contribution in its own right a catalogue of ESN metrics as the basis for future ESN analytics research.

Thirdly, to the body of *ESN research* more generally we contribute a unique set of ESN user roles derived from the data of a professional services firm, as a by-product of our method evaluation. While some initial work on ESN user roles exists (van Osch et al. 2016a; Frank et al. 2017; Cetto et al. 2018), to the best of our knowledge, our study is the first to explore user roles in ESN based on a systematic process and a comprehensive set of metrics. Moreover, we add specifically to the notion of KM-related roles in ESN (e.g., Cetto et al. 2018). Where the ESN is employed for knowledge work, digital traces can be understood as traces of knowledge interactions, such as discussing or sharing information. Being able to identify and distinguish different roles thus enables an understanding of how knowledge is created and continuously reproduced in the organization. In this respect, adding a new role identification method to the existing set of methods for the identification of informal networks and roles brings with it new opportunities for comparing different layers of organizational communication networks. Our case results provide a glimpse in the form of the *offline expert* or *sporadic user* roles, both of which were tagged in thanks or praise messages even though they did not warrant such mentions based on their ESN usage. We reasoned that these users enjoy a certain reputation or engage in certain activity in the organizational network outside the ESN which spills over into the ESN. The extent to which roles across different organizational networks (e.g. formal organization versus ESN) are congruent or divergent, is a promising area for future research.

Our method is further intended as a *practical contribution* for companies employing ESN. While many businesses have rolled out ESN, benefits are often unclear. For example, findings from role analysis might provide practitioners with insights about who is responsible for sharing knowledge in the network, guide staffing decisions and employee retention efforts, or identify key users as champions for organizational change efforts (see also Sect. 6.6).

8.2 Limitations and Future Work

Every study is circumscribed by certain design choices and the nature of its dataset and empirical access; ours is no exception. In the following we outline certain limitations, with a focus on the application, evaluation, and communication of EURIP, which point to opportunities for future research.

A number of limitations regarding the *application of EURIP* stem from the particular dataset and the empirical access given by the case company, as well as design choices we made for pragmatic reasons. Firstly, in our analysis we only considered active ESN users. Calculating the developed metrics for users with very few messages would not have led to meaningful measures. Moreover, we were interested specifically in users who actively contributed to the ESN. Yet, as with every ESN platform, the largest group of users will always be the group of non-active users that we did not include in our analysis. Such passive, or 'consuming', users do not post at all or very rarely, but may keep themselves informed using the ESN or even share information with co-workers outside of the ESN (Cranefield et al. 2015). While they are important, these users fell outside the bounds of our case study as the data did not facilitate the analysis of their behavior. Secondly, due to privacy considerations the content of the messages was not provided by the case company, so we had to rely on a limited number of keywords to develop and calculate metrics related to asking questions, thanking, and praising other users. Organizations who would use our method internally might not be bound by this limitation. Thirdly, we used one case only for demonstrating method feasibility and usefulness. Future case studies might be used to refine EURIP's structure, completeness, and further demonstrate validity. Given that the users we interviewed see themselves assuming several roles to varying degrees, the research team is planning to conduct fuzzy cluster analysis (e.g., Ferraro and Giordani 2015) which allows for a gradual assignment of users to different clusters. Finally, conducting the EURIP and making decisions based on the outcomes it produces can be considered as a people analytics application (Gal et al. 2017). Depending on particular legislation, such applications require the participation of a workers' council and the HR department (e.g., Fischbach et al. 2009). Whether or not the tracking and analysis of employee behavior on ESN is feasible, one needs to carefully think about how such information should be cross-referenced with other data, such as performance indicators (Gal et al. 2017) and the kinds of conclusions that can legitimately be drawn from it. Hence, there is a need for future research on how such ESN analytics can be done while respecting the privacy and confidentiality of individual employees.

Furthermore, we point out limitations regarding the *evaluation of EURIP*. Firstly, we were not able to carry out a broader evaluation, beyond selected interviews, utilizing survey methods to reach a greater number of employees. Secondly, both the interviewees and workshop participants could be considered ESN experts, and most of them were quite active on the ESN. While this increases confidence in the meaningfulness of the identified roles, it also means that the extent to which they identify with certain roles will be skewed, as we were unable to interview less active users. Finally, while workshop participants were enthusiastic about the potential uses of the identified roles for sense-making within the organization, due to lack of access we were not able to track actual use or application of our findings in subsequent decision-making processes.

References

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

Angeletou S, Rowe M, Alani H (2011) Modelling and analysis of user behaviour in online communities. In: Aroyo L, Welty C, Alani H et al (eds) The semantic web – ISWC 2011. Springer, Heidelberg, pp 35–50

Asamoah DA, Sharda R (2015) Adapting CRISP-DM process for social network analytics: application to healthcare. In: Proceedings of the 21st Americas conference on information systems, Puerto Rico

Backhaus K, Erichson B, Plinke W, Weiber R (2016) Multivariate Analysemethoden, 14th edn. Springer Gabler, Heidelberg

Behrendt S, Richter A, Riemer K (2014a) Conceptualisation of digital traces for the identification of informal networks in enterprise social networks. In: Proceedings of the 25th Australasian conference on information systems. Auckland

Behrendt S, Richter A, Trier M (2014b) Mixed methods analysis of enterprise social networks. Comput Netw 75(Part B):560–577

Berger K, Klier J, Klier M, Richter A (2014) "Who is Key…?" – Characterizing value adding users in enterprise social networks. In: Proceedings of the 22nd European Conference on Information Systems. Tel Aviv

Borgatti SP, Cross R (2003) A relational view of information seeking and learning in social networks. Manag Sci 49(4):432–445

Bortz J, Schuster C (2010) Statistik für Human- und Sozialwissenschaftler, 7th edn. Springer, Heidelberg

Brown JS, Duguid P (2001) Knowledge and organization: a social-practice perspective. Organ Sci 12(2):198–213

Bühner M, Ziegler M (2009) Statistik für Psychologen und Sozialwissenschaftler. Pearson, München

Burns MJ, Kotval XP (2013) Questions about questions: investigating how knowledge workers ask and answer questions. Bell Labs Tech J 17:43–61

Cetto A, Klier M, Richter A, Zolitschka JF (2018) "Thanks for sharing" – Identifying users' roles based on knowledge contribution in enterprise social networks. Comput Netw 135:275–288

Chan K, Liebowitz J (2006) The synergy of social network analysis and knowledge mapping: a case study. Int J Manag Decis Mak 7(1):19–35

Chan J, Hayes C, Daly E (2010) Decomposing discussion forums and boards using user roles. In: International AAAI conference on web and social media. AAAI Press, Washington, DC, pp 215–218

Chapman P, Clinton J, Kerber R, et al (2000) CRISP-DM 1.0: step-by-step data mining guide. https://www.the-modeling-agency.com/crisp-dm.pdf. Accessed 4 Nov 2019

Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36

Cios KJ, Swiniarski RW, Pedrycz W, Kurgan LA (2007) The knowledge discovery process. Data mining. Springer, Boston, pp 9–24

Cranefield J, Yoong P, Huff SL (2015) Rethinking lurking: invisible leading and following in a knowledge transfer ecosystem. J Assoc Inf Syst 16(4):213–247

Cross R, Prusak L (2002) The people who make organizations go – or stop effectiveness. Harv Bus Rev 80(6):104–112

Cross R, Borgatti SP, Parker A (2001a) Beyond answers: dimensions of the advice network. Soc Netw 23(3):215–235

Cross R, Parker A, Prusak L, Borgatti SP (2001b) Knowing what we know: supporting knowledge creation and sharing in social networks. Organ Dyn 30(2):100–120

Cross R, Borgatti SP, Parker A (2002) Making invisible work visible: using social network analysis to support strategic collaboration. Calif Manag Rev 44(2):25–46

Cross R, Kaše R, Kilduff M, King Z (2013) Bridging the gap between research and practice in organizational network analysis: a conversation between Rob Cross and Martin Kilduff. Hum Resour Manag 52(4):627–644

Csardi G, Nepusz T (2006) The igraph software package for complex network research. Int J Complex Syst Manusc 1695:1–9

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37

Ferraro MB, Giordani P (2015) A toolbox for fuzzy clustering using the R programming language. Fuzzy Sets Syst 279:1–16

Fischbach K, Schoder D, Gloor P (2009) Analyse informeller Kommunikationsnetzwerke am Beispiel einer Fallstudie. Wirtschaftsinformatik 51(2):164–174

Fischbach K, Schoder D, Putzke J, Gloor P (2010) Der Beitrag der Wirtschaftsinformatik zur Analyse und Gestaltung von informellen Netzwerken. In: Stegbauer C, Häußling R (eds) Handbuch Netzwerkforschung. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 679–686

Frank L, Gimpel H, Schmidt M, Schoch M (2017) Emergent user roles of a digital workplace: a network analysis based on trace data. In: Proceedings of the 38th international conference on information systems. Seoul

Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1(3):215–239

Friemel TN (2008) Netzwerkanalytische Methoden zur Identifizierung von Kommunikationsrollen. In: Stegbauer C (ed) Netzwerkanalyse und Netzwerktheorie. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 179–190

Füller J, Hutter K, Hautz J, Matzler K (2014) User roles and contributions in innovation-contest communities. J Manag Inf Syst 31(1):273–308

Gal U, Jensen TB, Stein M-K (2017) People analytics in the age of big data: an agenda for IS research. In: 38th international conference on information systems. Seoul

Gleave E, Welser H, Lento TM, Smith MA (2009) A conceptual and operational definition of "social role" in online community. In: 42nd Hawaii international conference on system sciences. IEEE, Big Island, pp 1–11

Graham T, Wright S (2014) Discursive equality and everyday talk online: the impact of "superparticipants". J Comput Commun 19(3):625–642

Hacker J (2017) Discovering knowledge actor roles in enterprise social networks – towards a better understanding of knowledge-in-practice. University of Erlangen-Nuremberg, Nuremberg

Hacker JV, Bodendorf F, Lorenz P (2016) A framework to analyze enterprise social network data. In: Atzmueller M, Oussena S,

Roth-Berghofer T (eds) Enterprise big data engineering, analytics, and management. IGI Global, Hershey, pp 84–107

Hacker J, Bernsmann R, Riemer K (2017a) Dimensions of user behaviour in enterprise social networks. In: Helms R, Cranefield J, van Reijsen J (eds) Social knowledge management in action: applications and challenges. Springer, Cham, pp 125–146

Hacker J, Bodendorf F, Lorenz P (2017b) A framework to identify knowledge actor roles in enterprise social networks. J Knowl Manag 21(4):817–838

Hacker JV, Johnson M, Saunders C, Thayer AL (2019) Trust in virtual teams: a multidisciplinary review and integration. Austral J Inf Syst

Hair JF, Black WC, Babin BJ, Anderson RE (2014) Multivariate data analysis, 7th edn. Pearson, Harlow

Hansen DL, Shneiderman B, Smith MA (2010) Visualizing threaded conversation networks: mining message boards and email lists for actionable insights. In: An A, Lingras P, Petty S, Huang R (eds) Active media technology SE – 7. Springer, Heidelberg, pp 47–62

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MISQ 28(1):75–105

Hevner AR, vom Brocke J, Maedche A (2019) Roles of digital innovation in design science research. Bus Inf Syst Eng 61(1):3–8

Holtzblatt L, Drury J, Weiss D (2013) Evaluating the uses and benefits of an enterprise social media platform. J Soc Media Organ 1(1):1–21

Jennex ME, Durcikova A (2013) Assessing knowledge loss risk. In: 46th Hawaii international conference on system sciences. IEEE, Wailea, pp 3478–3487

Junquero-Trabado V, Dominguez-Sal D (2012) Building a role search engine for social media. In: Proceedings of the 21st international conference companion on world wide web – WWW'12 Companion. ACM Press, New York, pp 1051–1060

Klier J, Klier M, Richter A, Wiesneth K (2017) Two sides of the same coin? – The effects of hierarchy inside and outside enterprise social networks. In: Proceedings of the 38th international conference on information systems. Seoul

Krackhardt D, Hanson JR (1993) Informal networks: the company behind the charts. Harv Bus Rev 71(4):104–111

Leonardi PM, Vaast E (2017) Social media and their affordances for organizing: a review and agenda for research. Acad Manag Ann 11(1):150–188

Leonardi PM, Huysman M, Steinfield C (2013) Enterprise social media: definition, history, and prospects for the study of social technologies in organizations. J Comput Commun 19(1):1–19

March ST, Smith GF (1995) Design and natural science research on information technology. Decis Support Syst 15(4):251–266

Martinez WL (2005) Exploratory data analysis with MATLAB. Chapman & Hall/CRC, Boca Raton

Morzy M (2009) On mining and social role discovery in internet forums. In: 2009 international workshop on social informatics. IEEE, Warsaw, pp 74–79

Newk-Fon Hey Tow W, Venable J, Dell P (2012) How organisations know what they know: a survey of knowledge identification methods among Australian organisations. In: Proceedings of the 23rd Australasian conference on information systems. Geelong

Oettl C, Berger T, Böhm M, et al (2018) Archetypes of enterprise social network users. In: Proceedings of the 51st Hawaii international conference on system sciences. Waikoloa Village, Big Island, pp 2036–2045

Parise S (2007) Knowledge management and human resource development: an application in social network analysis methods. Adv Dev Hum Resour 9(3):359–383

Parise S, Cross R, Davenport TH (2006) Strategies for preventing a knowledge-loss crisis. MIT Sloan Manag Rev 47(4):31–38

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24(3):45–77

R Core Team (2015) R: a language and environment for statistical computing. http://www.r-project.org/. Accessed 30 Mar 2019

Raiche G, Magis D (2015) Parallel analysis and non graphical solutions to the Cattell Scree Test. https://cran.r-project.org/web/packages/nFactors/nFactors.pdf. Accessed 31 Jan 2017

Richter A, Riemer K (2013) The contextual nature of enterprise social networking: a multi case study comparison. In: Proceedings of the 21st European conference on information systems, Utrecht

Riemer K, Scifleet P (2012) Enterprise social networking in knowledge-intensive work practices: a case study in a professional service firm. In: Proceedings of the 23rd Australasian conference on information systems, Geelong

Riemer K, Stieglitz S, Meske C (2015) From top to bottom. Bus Inf Syst Eng 57(3):197–212

Riemer K, Lee LL, Kjaer C, Haeffner A (2018) Metrics selection for group type identification in enterprise social network (ESN) analytics. In: Proceedings of the 29th Australasian conference on information systems, Sydney

Rowe M, Fernandez M, Angeletou S, Alani H (2013) Community analysis through semantic rules and role composition derivation. Web Semant Sci Serv Agents World Wide Web 18:31–47

Runkler TA (2015) Data mining. Springer, Wiesbaden

Sachs L (1978) Angewandte Statistik: Statistische Methoden und ihre Anwendungen. Springer, Heidelberg

SAS Institute (2012) SAS Enterprise Miner: SEMMA. https://web.archive.org/web/20120308165638/http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html/. Accessed 12 Dec 2017

Schendera CF (2010) Clusteranalyse mit SPSS: Mit Faktorenanalyse. Oldenbourg, München

Schubert P, Schwade F (2017) Social collaboration analytics for enterprise collaboration systems: providing business intelligence on collaboration activities. In: Proceedings of the 50th Hawaii international conference on system sciences. Waikoloa Village, pp 401–410

Schwade F, Schubert P (2019) Developing a user typology for the analysis of participation in enterprise collaboration systems. In: Proceedings of the 52nd Hawaii international conference on system sciences, Maui, pp 460–469

Steiny D, Oinas-Kukkonen H (2007) Network awareness: social network search, innovation and productivity in organisations. Int J Netw Virtual Organ 4(4):413

SWOOP Analytics (2019) SWOOP Personas – behavioural insights in social networks. https://www.swoopanalytics.com/personas/. Accessed 5 Feb 2019

Osch W van, Bulgurcu B, Kane GJ (2016a) Classifying enterprise social media users: a mixed-method study of organizational social media use. In: Proceedings of the 37th international conference on information systems. Dublin

Osch W van, Steinfield C, Zhao Y (2016b) Team boundary spanning through enterprise social media: exploring the effects of group-level diversity using a data science approach. In: 49th Hawaii international conference on system sciences. IEEE, Koloa, pp 2176–2185

Venable J, Pries-Heje J, Baskerville R (2012) A comprehensive framework for evaluation in design science research. In: Peffers K, Rothenberger M, Kuechler B (eds) Design science research in information systems. Advances in theory and practice. Springer, Heidelberg, pp 423–438

Venter J, de Waal A, Willers C (2007) Specializing CRISP-DM for evidence mining. In: Craiger P, Shenoi S (eds) Advances in digital forensics III: IFIP international conference on digital forensics. Springer, New York, pp 303–315

Viegas FB, Smith M (2004) Newsgroup crowds and authorlines: visualizing the activity of individuals in conversational cyber-spaces. In: 37th annual hawaii international conference on system sciences. IEEE, Big Island

von Krogh G (2012) How does social software change knowledge management? Toward a strategic research agenda. J Strateg Inf Syst 21(2):154–164

Welser HT, Gleave E, Barash V et al (2009) Whither the experts? Social affordances and the cultivation of experts in community Q&A systems. In: 2009 international conference on computational science and engineering. IEEE, Vancouver, pp 450–455

Whelan E (2011) It's who you know not what you know: a social network analysis approach to talent management. Eur J Int Manag 5(5):484