A Social Network Analysis Approach to People Analytics: Theoretical Underpinnings and Empirical Applications

Matthäus Zylka

University of Bamberg

Bamberg 2024

Diese Arbeit hat der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-Friedrich-Universität als Dissertation vorgelegen.

Erstgutachter: Prof. Dr. Kai Fischbach
Zweitgutachter: Prof. Dr. Peter Gloor
Mitglied der Promotionskommission: Prof. Dr. Tim Weitzel

Tag der Disputation: 11.06.2024

URN: urn:nbn:de:bvb:473-irb-959733
DOI: https://doi.org/10.20378/irb-95973

Dedicated to Mirabell, Elisa, Felizia & Jakob.

Table of Contents

Acknowledgments

Writing this dissertation has been one of the biggest challenges I have ever faced, and reaching the end of it feels almost unimaginable. When I left academia in 2017 and began working full time in consulting and software development, I foolishly thought I could juggle a job, family, and dissertation studies at the same time. It is unbelievable that I was able to complete this dissertation given the sacrifices it required, and it would not have been possible without the support of many people. I am grateful to all those who helped and, especially, encouraged me throughout the process.

Among them, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Kai Fischbach, for his unwavering support throughout the entire process of completing my PhD. His experience, mentorship, and encouragement have been invaluable to me and have greatly contributed to the success of this work. I would also like to thank the other members of my committee, Prof. Dr. Peter Gloor and Prof. Dr. Tim Weitzel. Thanks to Peter for taking the role as my second assessor. His good ideas, honest opinions, and research opportunities he offered me in the past years have been invaluable in my personal and professional growth. I am extremely grateful for his guidance and support. Thanks as well to my colleague and friend Prof. Dr. Oliver Posegga for his invaluable feedback to all my research projects, persistent pushing, and finally confronted me to *finish*.

Special thanks go to my wife, who demonstrated a lot of patience and at times provided the pressure I needed that pushed me forward. And I must thank my parents, who always supported and encouraged me to complete this dissertation, and who patiently listened to all my issues and doubts over the past few years.

Preamble

This work is written as a publication-based dissertation consisting of six peer-reviewed and published articles. All six articles, which are listed in Table i, are included in their accepted versions.

No.	Article	Status	Previous related research
I	Schreiber, R. R., Zylka, M. P. (2020). Social Network Analysis in Software Development Projects: A Systematic Literature Review.	Published in *International Journal of Software Engineering and Knowledge Engineering,* Vol. 30, No. 3, pp. 321-361	n/a
II	Zylka, M. P. & Fischbach, K. (2017). Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda.	Published in *ACM SIGMIS Database,* Vol.48, No. 2, pp. 52-78	Zylka, M. P. (2016). Putting the Consequences of IT Turnover on the Map: A Review and Call for Research. In: *Proceedings of the 2016 ACM SIGMIS Conference on Computers and People Research*, Best Paper Nominee. ACM, New York, NY, USA
III	Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K. & Schoder, D. (2015). Business Model Development, Founders' Social Capital, and the Success of Early-Stage Internet Startups: A Mixed-Method Study.	Published in *Information Systems Journal (ISJ)* Vol. 26, No. 5, pp.421-449	Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., Schoder, D. (2013). The Impact of Structural Embeddedness on Funding Success of Early-Stage Web Startups. In: *DIBME 2013 Pre-ECIS Workshop on the Digitization in Business Models and Entrepreneurship*, Utrecht, The Netherlands. Zylka, M. P., Spiegel, O., Fischbach, K. (2013). Employee Turnover in the IT Industry and its Impact on Social Capital and Success of Startup Companies. In: *XXXIII Sunbelt Social Networks Conference of the International Network for Social Network Analysis (INSNA)*, Hamburg
IV	Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Fischbach, K., Schlagwein, D., & Schoder, D. (2014). Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors.	Published in *Proceedings of the Academy of Management (AOM) Annual Meeting*	n/a
V	De Oliviera, J. M., Zylka, M. P & Gloor, P. A., Joshi T. (2019). Mirror, Mirror on the Wall, Who is Leaving of Them All - Predictions for Employee Turnover with Gated Recurrent Neural Networks.	Published in Springer Series *Studies on Entrepreneurship, Structural Change and Industrial Dynamics,* pp. 43-59	Zylka M.P., de Oliveira J.M., Gloor P.A., Joshi T. & Tickoo, P. (2018) Mirror, Mirror on the Wall, Who Is Leaving of Them All, In: *International Conference on Collaborative Innovation Networks Proceedings*, Suzhou, China.
VI	Gloor, P. A., Zylka, M. P., Fronzetti Colladon, A., Makai, M. (2022). 'Entanglement' – A new dynamic metric to measure team flow.	Published in *Social Networks,* Vol. 70, July, pp.100-111	n/a

Table i. Overview of articles included in this dissertation.

Abstract

The use of people analytics and social network analysis have grown in popularity in recent years as a way of gaining insights into organizational behavior. Social network analysis can be useful in people analytics the context because it can help identify key influencers and leaders within a group, as well as reveal patterns of communication and collaboration. Understanding these patterns, organizations can make informed decisions about how to structure and manage their teams and how best to support employee development. Social network analysis also helps people analytics to identify potential sources of conflict within an organization and bottlenecks in the flow of information, which in turn can aid in improving communication and productivity.

There is, however, still much to learn about the potential of social network analysis in a people analytics context. This thesis aims to contribute to this knowledge by examining the potential of using people analytics in conjunction with social network analysis as a way to understand and analyze relational data better, particularly communication and interaction data. To guide this investigation, the thesis begins with a narrative review of the theories and concepts that form the foundations of social network analysis, organizational network analysis, and people analytics, the overarching topic of this dissertation. This review underlies the development of a taxonomy for categorizing social network analysis studies in work-related contexts. The research conducted as part of this thesis is structured around one main research question and four subordinate research questions, all of which focus the understanding and application of social network analysis in a people analytics context. The findings from this research are presented in the six papers, which together form the core of the dissertation.

German Summary (Zusammenfassung)

Die vorliegende Arbeit befasst sich mit der Analyse von personenbezogenen Daten im Kontext von People Analytics. Dabei wird insbesondere den Kommunikations- und Interaktionsdaten von Mitarbeitenden Beachtung geschenkt. Im Mittelpunkt stehen Erkenntnisse, die durch Methoden der Netzwerkanalyse sowohl in interpersonellen als auch in interorganisationalen Netzwerken erlangt werden. In einem einleitenden Rahmenpapier beleuchte ich die theoretischen Grundlagen der Netzwerkanalyse, deren Anwendung in Organisationen und ihre Bedeutung in People Analytics. Weiterhin wird eine umfassende Literaturübersicht genutzt, um traditionelle Analyseansätze personenbezogener Daten mit relationalen Analysen in Unternehmen zu vergleichen. Die aktuellen relationalen Analysen in Unternehmen unterscheiden sich vor allem durch die zunehmenden Datenmengen (z.B. Kommunikationsdaten von Mitarbeitenden) und die jüngsten technologischen Fortschritte bei Analysemethoden und -verfahren (z.B. *maschinelles Lernen* und *neuronale Netze*). Aus diesem Vergleich und seinen Auswirkungen auf die Anwendung von Methoden und Theorien der Netzwerkanalyse leitet sich die zentrale Forschungsfrage der vorliegenden Arbeit ab:

Welchen Stellenwert hat die Analyse interpersoneller und interorganisationaler Netzwerke in People Analytics und welche Methoden haben einen entscheidungsunterstützenden Wert?

In diesem Rahmenpapier werden die Besonderheiten von People Analytics näher betrachtet und vor allem im Zusammenhang mit Methoden und Theorien der Netzwerkanalyse dargestellt, die eine zentrale Rolle in der interdisziplinären Forschung in Arbeitskontexten spielen. Auf der Grundlage dieser Untersuchung werden vier Unterforschungsfragen formuliert:

(1) Welche Rolle spielt die Analyse sozialer Netzwerke im Projektmanagement?

(2) Welche negativen und positiven Konsequenzen hat Mitarbeiterfluktuation für Mitarbeitende und Unternehmen?

(3) Welchen Erkenntnisgewinn haben Analysen von interpersonalen und interorganisationalen Netzwerken auf Basis von Mitarbeiterfluktuationen für People Analytics?

(4) Wie können Interaktions- und Kommunikationsdaten genutzt werden, um Vorhersagen über unternehmensrelevante Verhaltensmuster der Mitarbeitenden zu treffen?

Die Antworten auf diese Forschungsfragen präsentiere ich in sechs wissenschaftlichen Artikeln innerhalb dieser Dissertation. Im Rahmenpapier wird eine übergeordnete Taxonomie präsentiert, die die individuellen Beiträge den verschiedenen Kontexten der Analyse sozialer Netzwerke und People Analytics zuordnet. Zwei der sechs Artikel sind Literaturanalysen. Eine Literaturanalyse befasst sich mit der Identifikation von positiven und negativen Auswirkungen von Mitarbeiterfluktuation auf Mitarbeitende und Unternehmen. Die zweite Literaturanalyse stellt den aktuellen Forschungsstand der Anwendung der Analyse sozialer Netzwerke im organisationalen Kontext des Projektmanagements dar. Die verbleibenden vier Artikel sind anwendungsbezogene Studien, wobei zwei sich mit der Analyse von interorganisationalen und interpersonellen Netzwerken befasst, die auf Basis von Mitarbeiterfluktuations-Daten erstellt werden und die anderen beiden Studien sich auf die Vorhersage von unternehmensrelevanten Verhaltensmustern (u.a. Mitarbeiterfluktuation) konzentrieren.

Artikel I, der sich auf die erste Forschungsfrage bezieht, bietet mittels einer systematischen Literaturanalyse Einblick in den aktuellen Stand der Forschung und Anwendung der Netzwerkanalyse innerhalb von Unternehmen mit dem Fokus des Projektmanagements. Eine zentrale Erkenntnis der Studie ist, dass die Netzwerkanalyse in organisationalen Kontexten zur Identifizierung von Schwächen in der Kommunikation und Interaktion in Projektteams und zur Leistungsbeurteilung von Mitarbeitenden hilfreich sein kann und die Ursprünge negativer verhaltensbasierter Konsequenzen identifizieren kann.

Die Artikel II entfällt auf die zweite Forschungsfrage. Artikel II befasst sich mit den kognitiven, verhaltens- und leistungsorientierten Konsequenzen von Mitarbeiterfluktuationen und untersucht den aktuellen Forschungsstand zu diesem Thema. Eine Erkenntnis dieser Studie ist, dass es einen Mangel an Forschung zu positiven Konsequenzen von Mitarbeiterfluktuation auf die ehemaligen

Arbeitgeber/-innen und Kollegen/Kolleginnen als auch das Fehlen von Forschungsaktivitäten zu negativen Konsequenzen bei dem/r neuen Arbeitgeber/-in gibt. Die Studie gibt Anwendungshinweise und Forschungsrichtungen für zukünftige People Analytics Studien auf, sowohl in Bezug auf die bestehenden Forschungslücken als auch die untersuchten Aspekte in der Literatur.

Artikel III und IV beantworten die dritte Forschungsfrage. Artikel III untersucht mittels eines Mixed-Method-Ansatzes die Einflüsse von Sozialkapital von Unternehmensgründern auf den Unternehmenserfolg. Im qualitativen ersten Teil dieser Mixed-Methods-Studie zeigen 17 Experteninterviews, dass das Sozialkapital von Unternehmensgründern als entscheidender Faktor für die Entwicklung des Geschäftsmodells und letztlich für den Erfolg ihrer Startups angesehen werden. Im quantitativen zweiten Teil der Studie wird diese Behauptung auf der Grundlage einer Analyse der Mitarbeiterfluktuationen von 70 Internet-Startups und ihren 145 Gründern überprüft. Die Studie zeigt starke Belege für die entscheidende Bedeutung des Sozialkapitals der Gründer/-innen für den Erfolg von Internet-Startups auf. Artikel IV präsentiert ein Verfahren zur Erstellung von Industrieklassifikationen auf Basis von Mitarbeiterfluktuationen. In der vorgestellten Studie werden Community-Detection-Verfahren der Netzwerkanalyse auf einem gesammelten Datensatz angewendet, der interorganisationale Mitarbeiterfluktuationen dokumentiert, um Unternehmen gemäß des Verhaltens ihrer Mitarbeitenden, das durch Personaldaten abgebildet wird, zu klassifizieren. Der Beitrag demonstriert nicht nur das Potential des Verfahrens und ordnet es ein, sondern bietet auch empirische Einblicke in die Struktur der IT-Branche.

Artikel V und VI entfallen auf die vierte Forschungsfrage, die sich mit der Vorhersagekraft von neuartigen methodischen Ansätzen der Analyse sozialer Netzwerke befasst. Artikel V adressiert die Forderung nach innovativen Methoden zum prädiktiven und präskriptiven People Analytics. Die bisherigen Ansätze zur u.a. Vorhersage von Mitarbeiterfluktuationen durch klassische statistische Methoden ist nur gering bis mäßig. Um dieses Manko zu beheben, führen wir ein Deep-Learning-Experiment zur Vorhersage der Mitarbeiterfluktuation eines international agierenden Beratungsunternehmens durch. Basierend auf einem einzigartigen Datensatz, der eine 12-monatige Zeitreihe der E-Mail-Kommunikation von 3952 Managern enthält, zeigen wir, dass neuronale Netzwerke auf Basis von Metriken der Netzwerkanalyse das Potential haben, klassische Vorhersagemodelle zu übertreffen. Artikel VI führt ein neues Maß namens *Entanglement* zur Analyse sozialer Netzwerke ein. Entanglement ist eine neuartige Metrik zur Messung der Synchronität der Kommunikation zwischen Teammitgliedern. Dieses Maß berechnet die euklidische Distanz zwischen den Zeitreihen ausgewählter Zentralitätsmetriken der Teammitglieder. Wir validieren die Metrik in der Studie anhand von vier Fallstudien. Die erste Fallstudie verwendet die Vernetzung von 11 medizinischen Innovationsteams, um die Teamleistung und das Lernverhalten der Teammitgliedern vorherzusagen. In der zweiten Fallstudie wird die E-Mail-Kommunikation von 113 Führungskräften eines internationalen Dienstleistungsunternehmens untersucht, um die Mitarbeiterfluktuation aufgrund mangelnder Vernetzung im Kommunikationsnetzwerk vorherzusagen. In der dritten Fallstudie wird die individuelle Mitarbeiterleistung von 81 Managern analysiert. Im Rahmen der vierten Fallstudie sagen wir die Leistung von 13 kundenorientierten Teams in einem großen internationalen Unternehmen voraus, indem wir die Vernetzung durch E-Mail-Interaktion mit der Zufriedenheit ihrer Kunden/Kundinnen vergleichen, die anhand des *Net Promoter Score* gemessen wird. Studie VI stellt fest, dass es sich bei der vorgeschlagenen Metrik um einen neuen und vielseitigen Indikator für People Analytics handelt, der die bisher wenig genutzte zeitliche bzw. dynamische Dimension organisationaler Netzwerke analysiert und als leistungsfähiger Prädiktor für Mitarbeiter- und Teamleistung, Mitarbeiterfluktuation und Kundenzufriedenheit eingesetzt werden könnte.

Von den in der Dissertation zusammengefassten Artikeln wurden vier Artikel (I, II, III, VI) in einschlägigen Journalen (*International Journal of Software Engineering and Knowledge Engineering, ACM SIGMIS Database, Information Systems Journal* und *Social Networks)* veröffentlicht. Artikel IV wurde in einem Tagungsband (*Proceedings of the Academy of Management (AOM) Annual Meeting)* und Artikel V in einem Sammelband (*Springer Series*: *Studies on Entrepreneurship, Structural Change and Industrial Dynamics*) veröffentlicht.

1

Introduction

Over the past few decades, there has been a considerable increase in interdisciplinary research attention to employee behavior that uses data from human resources information systems and other digital work-related data sources such as email archives or communication platforms in general. There have been a significant number of applications within organizations themselves. The growing volume and pervasiveness of employee-related data is one of the main reasons behind this phenomenon. Interaction and communication data, which can indicate the work performance of employees and thus of organizations as a whole, have been a significant part of these data. A survey conducted by McKinsey & Company (2012) found that employees spend 28% of their workweek, on average, reading, writing, and responding to emails, which gives a sense of how much communication data is generated in organizations each day.

The analysis of interaction and communication data raises expectations about the technology and methodologies needed. While technological barriers (e.g., missing analytical power) can be solved by increased investments in information technology, matching methodologies must be tailored to the analytical goals of the organization and its data analysts.

Social network analysis is a set of theories and methods (Borgatti & Halgin, 2011) that provides an appropriate toolbox that can be tailored for analyzing communication and interaction among individuals and groups. Its procedures (mathematical and graphical techniques) use indications of relatedness among entities to represent social structures in a compact and systematic manner (Butts, 2009).

Organizational network analysis is an application of social network analysis in work-related contexts. It comprises a set of theories and methods for mapping and measuring relationships between employees, teams, and organizations with the resources, knowledge, and tasks used to perform work (Borgatti & Foster, 2003). It draws upon theories of organizational theory and network theory (Borgatti & Halgin, 2011) to produce models representing complex interpersonal and interorganizational interactions that would be impossible to describe without relational concepts. The resulting insights can help managers understand critical performance factors such as how information diffuses among individuals and influences the speed, quality, and accuracy of organizational decisions (Galbraith, 1974). Further, these insights can reveal where resources are inadequate for employees to perform their work. Thus, social network analysis and organizational network analysis can support planning for and justifying the allocation of resources and aid decisionmaking by revealing links between information networks and process performance.

People analytics, also known as human resources analytics, is a data-driven approach to managing people and processes in organizations that supports decisionmaking (Marler & Boudreau, 2017). It involves the use of human resources data, especially employee demographic data (work history, gender, tenure, etc.), and analytics to understand and predict the behavior and performance of employees. People analytics is seen as the key component of many organizations' human resources strategies, with the use of data and analytics playing a critical role in driving business decisions and improving the employee experience.

Some companies have implemented people analytics and have used it for operational reporting and decision making, but many others still struggle to do so. Those that have focused on people analytics have often found it difficult to transition from basic reporting to more advanced statistical analysis and the identification of causal dimensions as a way to deliver actionable solutions (Minbaeva, 2017; Schiemann et al., 2018). Here is where social network analysis can help people analytics by providing a theoretical lens and methods to analyze and understand the human resources data from a relational perspective (Leonardi & Contractor, 2018). For example, communication and interaction data could be analyzed to identify key influencers within an organization, study the impact of communication network structures on team performance, or examine the impact of social network positions on individual career outcomes (Cross & Parker, 2004; Kilduff & Tsai, 2003). There is an extensive and rapidly growing body of scientific literature on the application of social network analysis in organizational contexts (Borgatti & Foster, 2003; Brass, 2003, 2022; H. Chen et al., 2022), as well as case studies of real-world applications in organizations (e.g., Deloitte, 2020).

Thus, the goal of this thesis is to provide an overview of the role of social network analysis in people analytics and explore how social network analysis can be used to improve the predictive power of people analytics. Further, it discusses and provides various analytical approaches that are rooted in social network analysis.

This work poses the four main research question:

(1) What role does social network analysis play in managing projects?

(2) What are the potential positive and negative consequences of employee turnover for employees and companies?

(3) How can social network analysis be applied to assess the consequences of employee turnover for interpersonal and interorganizational networks relevant for people analytics?

(4) How can social network analysis be applied to predict company-relevant employee behavior patterns for people analytics?

Each of the six research papers presented herein contribute in some way to these four questions. Two of the papers are literature reviews and four are applications of social network analysis in a people analytics context.

This thesis is structured as follows: The next section provides a theoretical overview on social network analysis and its application within organizations. It is followed by a section dedicated to background on people analytics. Based on this theoretical background, I create a classification framework aimed at categorizing the six papers comprising this thesis, show their similarities and differences with respect to diverse constructs of social network analysis and people analytics. The framework section is followed by a discussion of how the papers contribute to the four research questions of this thesis, based on the classification framework. The limitations of this research are discussed, and the dissertation concludes with a summary of the contributions. The papers comprise the Appendix.

2

Theoretical Foundations

This section explores the topic of social network analysis and its application in analyzing social relationships and networks in organizations. It also delves into the concept of people analytics, discussing its origins and role in human resources management. Specifically, it examines how social network analysis can be utilized in people analytics to optimize human resources practices and enhance an organization's overall performance and success.

Social Network Analysis

The field of social network analysis has a long and rich history, with roots dating back to Georg Simmel's concept of the intersection of social circles in the early 1900s (Simmel, 1908). One of the earliest and most influential pioneers of *modern* social network analysis (Freeman, 2004) was sociologist Jacob Moreno, who developed a methodology called *sociometry* for studying the patterns of social interactions and relationships within groups (Moreno, 1937). Moreno's work laid the foundation for many of the theories, concepts, and techniques that are now central to social network analysis, including the use of graphs and matrices to represent social networks and the analysis of centrality (Carrington et al., 2005; Freeman, 1977), as well as the relative importance of individuals within a network.

Social network analysis continued to evolve as more researchers began to use it to study social interactions in a variety of contexts. Two notable examples of the application of social network analysis are the studies of sociologist Mark Granovetter and Ronald S. Burt. Granovetter's theory on the *strength of weak ties* (1973)— the importance of more distant connections in social networks—has become a foundation of social network analysis. Similarly, Burt's introduction of the theory of *structural holes* (1992)—gaps between individuals who have complementary sources of information—has also become a crucial theory within the field.

The theoretical origins of social network analysis focus on the role of social networks in interpersonal relationships and social structures and can be used to understand the mechanisms that shape the beliefs and behaviors of individuals, groups, and organizations (Kilduff & Tsai, 2003). This includes small and large groups, departments within an organization, and the organization as a whole (Borgatti & Foster, 2003). By mapping and analyzing the relationships between people, researchers can study group dynamics (Borgo, 2006; Gloor et al., 2012; Hasan, 2012) and how individuals leverage their connections to achieve desired outcomes (Coleman, 1988; Corredoira & Rosenkopf, 2010). The validity of the theoretical foundation is often assessed through the use of social network analysis (Howison et al., 2011).

Social network analysis is a multifaceted field that encompasses a wide range of methods and techniques (Carrington et al., 2005; Howison et al., 2011) and has been applied in various domains (Borgatti et al., 2009; Girvan & Newman, 2002; Lazer et al., 2009; Oinas-Kukkonen et al., 2010). The fundamental method of social network analysis is to model social relationships and interactions between individuals as graphs comprised of nodes representing individual actors and edges representing their relationships (Butts, 2009). Social network analysis has multiple objectives, including understanding the structure and patterns of relationships within the network (Barabási, 2002), examining the impact of social relationships on individual behavior and outcomes (Granovetter, 1973), analyzing the flow of information and resources within the network (Burt, 1997), predicting future events and outcomes based on network structures (Lazer et al., 2009), identifying key drivers of social change within the network (Moody & White, 2003), evaluating the effectiveness of interventions and strategies aimed at modifying the network (Centola & Macy, 2007), and developing strategies for managing and optimizing the network to achieve specific goals (Kleinberg, 2000).

As mentioned above, social network analysis can be applied in organizational studies. This application of social network analysis is specified as *organizational network analysis*. In organizational network analysis, researchers use network theory and organizational behavior theory and apply social network analysis methods to understand the relationships between employees (Brass, 1981), organizations (Burt, 1992), and other non-human elements within an organization and to investigate their effects on organizational outcomes such as job attitudes and performance (Raz et al., 2007; Uzzi, 1996).

An organizational network is a system of informal cooperation between multiple interdependent actors, which is established through the division of labor and the creation of value through synergy. The functioning of an organization is based primarily on interactions and the exchange of knowledge and information between people working together (Nonaka, 1994). These interactions create relationships, and the form and intensity of these relationships shape *micro-level* and *macro-level* networks within the organization (Granovetter, 1973; Mead, 1934). Micro-level networks (often called *interpersonal* or *intraorganizational networks*) focus on relationships between employees with specific attributes (e.g., years of tenure) or memberships (e.g., team or department), whereas macro-level networks (or *interorganizational networks*) mainly concern relations between organizations (Galaskiewicz & Wasserman, 1994).

The internal dynamics of an organization also influence the network structure of relationships between its elements (H. Chen et al., 2022). Organizations, teams, and even individual employees are in a constant state of flow. The flow of information, knowledge, resources, and tasks is inherently dynamic and provides valuable insights into the patterns of interactions between people and departments in the organization. This understanding is what shifted the research focus from the analysis of network structure to an increasing exploration of network dynamics (Brass, 2022; H. Chen et al., 2022). Network dynamics examine networks whether the ties are relational states (e.g., collegiality) or events (e.g., sent messages) that occur over relatively short time scales (Borgatti & Halgin, 2011). Examples of events that have been studied in micro-level networks include physical interactions between people tracked by sociometric badges (Fischbach et al., 2008) and emails exchanged between several employees (e.g., Gloor et al., 2014).

Studies that focus on events often aim to understand their sequence and timing and how they relate to the flow of information between individuals or organizations (Borgatti & Halgin, 2011). Event-based data are a valuable source of information because they allow for tracking and analyzing specific events or actions that have occurred within the organization (H. Chen et al., 2022). By analyzing these data over time, organizations can gain insights into how their processes are functioning and identify areas for improvement. Time-stamped event data, for instance, can be particularly useful in understanding how communication patterns within an organization change over time (H. Chen et al., 2022). This can help organizations understand the effectiveness of their communication channels and identify potential bottlenecks or inefficiencies in their processes (Scott, 2000). Event-based data can also be useful in identifying trends or patterns that may not be immediately obvious from other types of data (H. Chen et al., 2022). At the macro-level, such patterns can have a structural and temporal nature (Kitts et al., 2017). As mentioned above, gaining insights about performance and efficacy of employees and behavioral patterns are very important topics of organizational network analysis.

The relationship between performance and network has been studied quite extensively. The study of organizational performance on a micro level is an important application of network analysis (Borgatti & Foster, 2003). By analyzing communication patterns within an organization, it is possible to identify key influencers and decisionmakers and understand how information flows through the organization (Barabási, 2002). Sparrowe et al. (2001) found that individuals who were central in their workgroups' advice networks had higher levels of both in-role and extra-role performance compared to those who were not central in such networks. Conversely, those who were central in a hindrance network, which is a network of relationships with colleagues who impede task behaviors, had lower levels of performance in both their assigned duties and additional responsibilities (Sparrowe et al., 2001). These results suggest that being central in a group's advice-sharing network is associated with more positive evaluations of individual performance (Sparrowe et al., 2001).

The study by Soda et al. (2021) suggests that the ability of individuals to benefit from their network in terms of creative performance is influenced by the network's dynamic nature. Specifically, the study found that the positive effect of structural holes on individual creativity was greater when the network structure was dynamic, with the addition of new ties over time (Soda et al., 2021). This suggests that maintaining a dynamic network structure, with the incorporation of new connections and relationships, can be beneficial for individual creativity (Soda et al., 2021). Conversely, when the network structure becomes more static, with fewer changes in the ties that make up the structural holes, the benefits of these structural holes for creativity may disappear as a result of the increased inflexibility in thinking and coordinating (Soda et al., 2021).

Performance has been studied at a macro level in the context of patent ownership in the semiconductor industry (Podolny et al., 1996), pricing strategies (Podolny, 1993), and power relations (Baker, 1992) of investment banks and alliances (Chellappa & Saraf, 2010; Goerzen, 2007). Chellappa & Saraf (2010) suggest that a firm's relative structural position within its alliance network can be used as a proxy for its standards dominance and can be indicative of its performance. Rather than using structural measures designed for interpersonal networks, they propose a measure of relative firm prominence specifically for the business software network, where the benefits of alliances may still be obtained even through weakened indirect connections (Chellappa & Saraf, 2010).

At the micro level, the consequences that derive from behaviors have been studied. This includes employee turnover, a critical behavior that can have serious impacts on individuals and organizations (Staw, 1980; Zylka & Fischbach, 2017). In recent years, researchers have been interested in studying employee turnover from a relational perspective (Holtom et al., 2008), taking into account not only individual factors, but also organizational, environmental, and economic factors. There has also been growing interest in the role of employee social relationships in predicting employee turnover. This is based on the idea that having strong social connections at work (also known as social capital) can improve job satisfaction and reduce turnover (Gloor et al., 2017). For example, a study by Krackhardt and Porter (1986) found that turnover at a fast-food restaurant was related to groups of employees who had similar structural positions and communicated more with each other. Woehler et al. (2021) examined the impact of corporate mergers on employee personal networks and turnover intentions, using a combination of survey data and email communication data to analyze the behavior of employees at two firms before and after their merger. They found that employees who increased the number of connections they had with individuals from the other company during the merger process were less likely to leave the firm (Woehler et al., 2021); the researchers attributed this to the fact that these employees were able to gain valuable information and support from their connections in the other company, which helped them adjust to the changes brought about by the merger.

In macro-level network studies, researchers have been interested in corporate behavior such as rivalry and competitive actions (Hofer et al., 2022; Porac et al., 1995) and collaborations (Park et al., 2011). However, many macro-level studies link behaviors such as competitive actions to performance and do not focus on the actual behavior (e.g., Hofer et al., 2022).

Over the years, a wide range of measures have been developed to study how the patterns of connections among individuals and groups in a social network can affect various outcomes such as performance and behavior. These measures, often referred to as structural measures, describe the structure or configuration of relationships within a network. Structural measures can be further divided into point measures, which describe an individual's position within a network, and whole network measures, which describe the configuration of the entire network or a subnetwork within it (Brass, 2022). Point measures of centrality, for example, can be used to quantify an individual's importance within a network in terms of that individual's number of connections (degree centrality), the proximity of their connections to other individuals in the network (closeness centrality), and their ability to connect other individuals in the network through the shortest paths between them (betweenness centrality) (Brass, 2022; Freeman, 1977).

Studying the presence and size of structural holes can provide insights into how information flows within an organization (Burt, 1992). Density, as a whole network measure, reflects the extent to which individuals or groups in a network are connected. A dense network is one in which most individuals are connected to each other, while a sparse network is one in which there are fewer connections (Scott, 2000).

Clustering coefficient reflects the extent to which individuals or groups in a network tend to cluster together (Watts & Strogatz, 1998). A high clustering coefficient indicates that individuals are more likely to be connected to each other through multiple pathways.

Assortativity reflects the extent to which individuals or groups in a network tend to be connected to others who are similar to them (Newman, 2002). For example, employees with similar job titles or departments may be more likely to be connected to each other.

The section that follows first presents the roots of people analytics and then explains subsequently how social network analysis is related to people analytics and can provide novel analytical approaches to people analytics by using the various measures and concepts spelled out above.

People Analytics

People analytics is a data-driven approach to managing people at work. Business leaders can make decisions about their employees based on extensive analysis of data rather than using the traditional methods of personal relationships, decisionmaking based on experience, and risk avoidance.

The development of people analytics can be traced back to the 1960s, when the research stream of organizational behavior studies began to emerge (Argyris, 1960). Researchers and practitioners began to explore the use of data and analytics to understand the relationship between the formal structure of an organization and individual employee behavior (Argyris, 1960) and factors that drive organizational performance (Etzioni, 1960). This included the use of statistical analysis, survey research, and other methods to collect and analyze data on employee attitudes, behaviors, and outcomes (Argyris, 1960; Presthus, 1962). Later, the field of analytics in human resources management continued to evolve, with a focus on using data and analytics to support decisionmaking in human resources departments (DeSanctis, 1986). Human resources information systems and other technologies allowed organizations to collect and analyze data on employee demographics, performance, and other factors (Tannenbaum, 1990). The rise of "Big Data" and advances in analytics technologies such as *machine learning* and *artificial intelligence* have further expanded people analytics, but have also added complexity to the data analysis (Angrave et al., 2016). Organizations have since begun to use data and analytics to support a wider range of human resources functions, including recruitment (Laumer et al., 2022), workforce planning, and employee engagement (Isson & Harriott, 2016).

Organizations are using people analytics to gain a competitive advantage by defining metrics for human capital investment, analyzing data and considering potential scenarios for future workforce needs, and improving talent retention (Isson & Harriott, 2016). This is known as human resources analytics in the academic literature (Marler & Boudreau, 2017). Research has shown that there are complementarities between information technology, performance pay, and human resources analytics practices (Aral et al., 2012). However, human resources analytics will not bring about desired transformations unless several issues are addressed and better methods and approaches are developed (Angrave et al., 2016). These include maximizing value from human resources data, measuring and modeling the impact of human capital inputs on strategy, and overcoming limitations and a lack of understanding within human resource departments and analytics teams. Measuring and understanding employee engagement has emerged as an important challenge and opportunity in human resources analytics.

People analytics and human resources analytics are closely related; both involve using data and analytical techniques to understand and improve the effectiveness of human resources policies and practices. There are, however, some key differences. One of the main differences between people analytics and human resources analytics is the focus; people analytics is a broader term that encompasses a range of analytical techniques and approaches that can be used to understand and improve the performance and wellbeing of employees, whereas human resources analytics is more narrowly focused on using data and analysis to understand and improve human resources policies and practices (Huselid, 2018). Another key difference concerns scope: people analytics encompasses a wide range of data sources and analytical techniques, and can be applied to a variety of human resources-related issues, such as workforce planning, talent management, and employee engagement; human resources analytics typically focuses more on specific human resources processes, such as recruitment and performance management (Tursunbayeva et al., 2018).

The following understanding of people analytics guides this thesis:

People analytics is a field within human resource management that focuses on using data analytics and visualization tools to understand workforce dynamics, human capital, and individual and team performance. It involves the use of information technologies and data analytics to generate insights that can be used to make strategic decisions and improve organizational effectiveness, efficiency, and outcomes, and to enhance the employee experience.

People analytics can also be considered as a part of the broader literature on information technology and human resources management. Stone et al. (2015) highlight the significant impact of information technology on human resources processes, but also point out the limitations of existing human resources systems with respect to attracting and retaining employees. Technology has transformed the nature of work and the management of human resources processes through the use of e-human resources management, such as e-recruiting, e-learning and training, employee performance management, and e-compensation systems. The authors argue that the main goal of e-human resources management should be effectiveness in achieving human resources objectives, rather than simply the efficiency of processes (Stone et al., 2015).

Wirtky et al. (2016) conducted a systematic review of the e-human resources management literature since 1990, covering major human resources management functions including planning, internal staffing, external recruiting, development, motivation, and administration. The research examined the role of information technology in human resources management functions and the implementation and adoption of human resources management systems, particularly in the context of the current "war for talent" that has emerged from labor market shortages (Wirtky et al., 2016). The authors pointed out opportunities for research in the areas of transformative human resources management capabilities and talent management, including identifying talent across various skills such as leadership and knowledge.

Tools for analytical forecasting and decision making in people analytics are relatively new and not as advanced as those used in other business areas such as marketing, sales, production planning, and finance (Schiemann et al., 2018). Many studies have noted that current data analysis approaches in this field are mostly limited to reactive, standardized, and/or historical reporting based on demand, as well as simple forecasts (Boudreau & Cascio, 2017; Minbaeva, 2017). As a result, many organizations are still trying to move beyond basic operational reporting and toward more advanced forms of analytics.

Analytics in general can be categorized as descriptive, predictive, or prescriptive (Davenport & Harris, 2007; Lustig et al., 2010). *Descriptive* people analytics refers to the use of data and analytics to describe and summarize existing human resources data, such as employee demographics and productivity; this involves using basic statistical and visualization techniques to understand the current state of human resources data, and to identify patterns and trends in this data (King, 2016). *Predictive* people analytics refers to the use of data and analytics to make predictions about future human resources trends and outcomes; this involves using advanced analytical techniques, such as machine learning and predictive modeling, to build models that can forecast human resources trends and outcomes based on past data (Schafheitle et al., 2020). Finally, *prescriptive* people analytics refers to the use of data and analytics to recommend actions and interventions to improve human resources and people management in organizations; this involves using advanced analytics to identify and evaluate potential human resources interventions, and to recommend the most effective and efficient interventions based on the data and analysis (Fernandez & Gallardo-Gallardo, 2021). Overall, these three categories—descriptive, predictive, and prescriptive people analytics—are in a continuum of different levels of maturity and capability in the use of data and analytics to support human resources and people management in organizations (Lustig et al., 2010). As organizations move from one stage to the next, they are able to use data and analytics in increasingly sophisticated and effective ways to improve the employee experience and drive organizational performance (Burton et al., 2020; Schafheitle et al., 2020).

The objectives of people analytics can be summarized as falling into two main categories: performance (Kryscynski et al., 2018; Sharma & Sharma, 2017; Wang & Cotton, 2018) and behavior (Chamorro-Premuzic & Bailie, 2020), indicating a diversification of people analytics and a growing focus on relationships caused by behavior and vice versa. They include performance assessment and development, employee lifetime value, onboarding and culture fit, engagement, and workforce planning (Tursunbayeva et al., 2018). A thematic analysis identified four categories: employee collaboration, diversity and inclusion, people risks, and inter-organizational relationships (Tursunbayeva et al., 2018).

Leonardi and Contractor (2018) call for focusing on relational data in order to overcome the low analytical power of current people analytics efforts. They argue that this relational perspective should replace traditional organizational design approaches (Leonardi & Contractor, 2018) because of its potential for analyzing interactions, affiliations, and the performance of groups, including the use of data on social and organizational networks. Their call for the relational perspective could be addressed by social network analysis. For example, people analytics could use social network analysis to identify employees with strong ties to external networks and/or customers and encourage them to share their insights and connections with their coworkers. Understanding the strength of weak ties can also assist a company in identifying potential candidates for job openings or to work on new projects. By analyzing the social networks of current employees, a people analytics team could identify individuals with a diverse range of connections who may be able to refer qualified candidates or bring new ideas to the table. Further, weak ties can also be used to promote workplace diversity and inclusion. An organization could broaden its network and create more diverse and inclusive opportunities for all of its employees by identifying and cultivating relationships with employees who have weak ties to underrepresented teams.

According to the structural holes theory, individuals or organizations that occupy structural holes, or gaps, in a social network are more likely to have access to valuable resources and opportunities (Burt, 1992). In the context of people analytics, this theory can be applied by analyzing the social networks of employees to identify individuals who occupy structural holes and may be able to facilitate connections and information flow between different parts of the organization or external networks. By identifying structural holes within an organization, people analytics teams can also identify opportunities to create new connections and facilitate the flow of information and resources. In addition, structural holes theory can be used to identify and address potential barriers to information flow within an organization, such as isolated groups (*silos*) or departments. Burt (2004) showed that between-groups brokers—employees who bridge structural holes between teams—are more likely to express ideas and are evaluated as more valuable. This specific position in a network tends to come with greater social capital than others (Burt, 2004).

As mentioned in the social network analysis section, social network analysis does not provide only a theory (Borgatti & Halgin, 2011), but also measures which help to identify specific positions, such as brokers in micro- and macro-level networks. For example, centrality measures are proxies for importance, prestige and power (Freeman, 1977). A high betweenness centrality for instance, could be used to find brokers in a network.[1]

Modularity, or the degree to which a network can be divided into distinct subgroups (Newman, 2006), can also be useful in people analytics, for several reasons. By analyzing the modularity of an organization's social network, data analysts could identify subgroups that are highly interconnected and may be more effective at collaboration and problem solving. A modularity-based analysis can also help identify subgroups that may be isolated from the rest of the organization (silos) and in need of stronger connections

[1] Betweenness centrality is only one variant for measuring brokerage. There are many other variants of shortest-path centrality algorithms that can be used for brokerage (Everett & Valente, 2016).

to facilitate information flow and collaboration. In addition, understanding the modular structure of an organization can inform the design and implementation of team-building and communication strategies that align with the natural patterns of interaction within the organization. The modularity measure can also be used to identify potential conflicts or bottlenecks within an organization, as well as opportunities to create new connections and facilitate the flow of information and resources between different parts of the organization.

In summary, social network analysis could improve people analytics in two ways. First, it provides a powerful toolset, with measures for relational analyses of organizational data such as communication and interaction data combined with demographic data of employees. Second, it can support descriptive and predictive people analytics by providing a theoretical lens, namely *network theory* (Borgatti & Halgin, 2011), which is helpful for the interpretation of the analysis results.

3

Categorization of Articles

Based on the insights gained from the theoretical background section, the overarching framework shown in Figure 1 was built. Although social network analysis and people analytics represent concurrent independent research streams, social network analysis can—as the figure suggests—help people analytics efforts by taking the relational lens of social network analysis into account. Further, people analytics benefit from *network theory* because it "refers to the mechanisms and processes that interact with network structures to yield certain outcomes for individuals and groups" (Borgatti & Halgin, 2011, p. 1); this matches with people analytics efforts aimed at understanding workforce dynamics and individual and team performance and generating insights that can be used to make strategic decisions and improve organizational effectiveness, efficiency, and outcomes. The two research streams encompass the six major concepts discussed in Section 2 and emphasize two major consequences: performance and behavior.

Article I investigates the applications of social network analysis in project management and provides an overview of current structural analysis efforts at the micro level, focusing on performance and behavior outcomes. Article II contributes to the fundamental understanding of the consequences of individual employee turnover. Article III investigates the influence of social capital on the funding success of early-stage startups, and demonstrates how social network analysis methods can be applied to micro- and macro-level networks and derive metrics for descriptive people analytics. Article IV uses data on employee turnover in an explorative approach to derive a novel industry classification system. Both Articles III and Article IV analyze state-based ties, since employee turnover is by definition not an event that occurs over a relatively short period of time scales. Articles V and VI analyze, from a microlevel perspective, event-based ties formed by communication data; both articles apply machine learning and neural networks. Article V focus on behavioral outcomes only, whereas Article VI also analyzes the performance of employees and employee behavior.

The articles are classified based on their level of analysis (micro- or macro-level network), whether the article applies a structural analysis of networks, and the network dynamics perspective on social ties adopted, that is, whether the ties are states-type ties (e.g., collegiality) or event-type ties (e.g., sent message) that occur over relatively short periods of time. The articles are further classified according to where they fall on the continuum of maturity and capability (descriptive, predictive, and prescriptive) for people analytics discussed in Section 2.

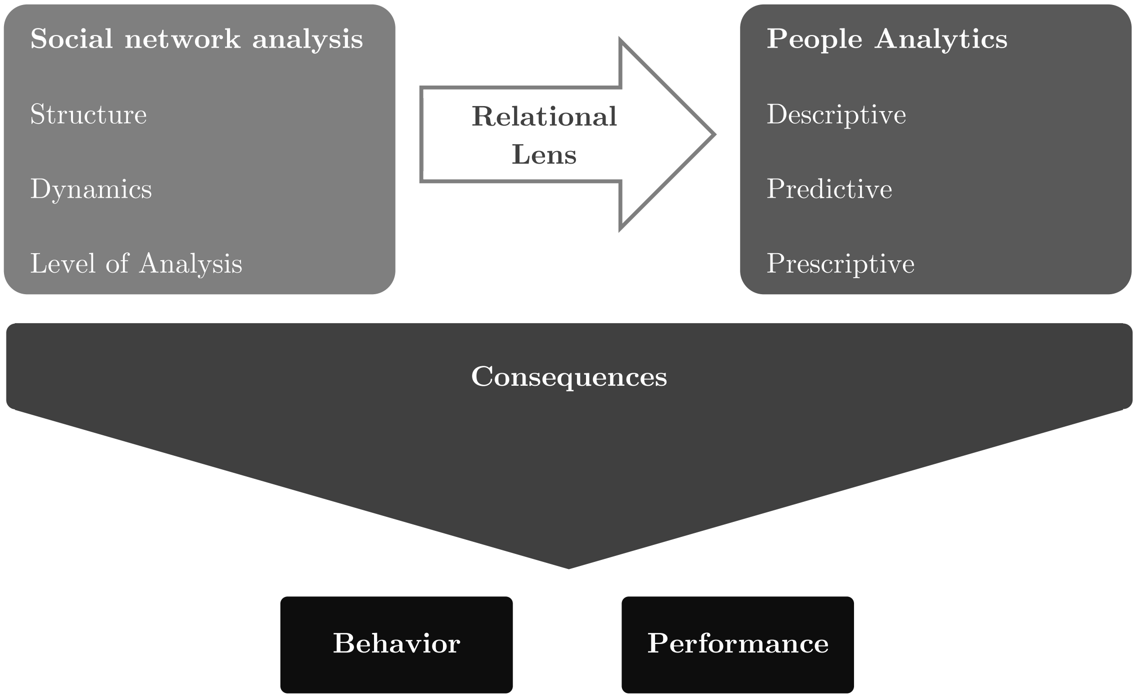


Figure 1. Overarching framework.

Table 1 shows the classification of the six articles within this overarching framework and summarizes each article's focus of the six research articles. Each of the table's columns correspond to a different aspect of social network analysis, people analytics, and consequences. The "Structure" column indicates whether the article provides insights about the structural analysis of network; "Dynamics" whether the article focuses on state-type or event-type ties; and the "Micro level" and "Macro level" columns indicate whether the article focuses interpersonal networks (on micro-level) or interorganizational networks (macro-level). The "Descriptive," "Predictive," and "Prescriptive" columns indicate the degree to which the article is focused on describing, predicting, or prescribing actions based on the analyzed data. The "Performance" column indicates whether the article focuses on the performance outcomes of the analysis; and "Behavior" whether the article focuses on the behavior or actions of individuals or organizations.

	Social Network Analysis					People Analytics			Consequences	
		Dynamics		Level						
No.	**Structure**	**State**	**Event**	**Micro**	**Macro**	**Descriptive**	**Predictive**	**Prescriptive**	**Performance**	**Behavior**
I	X			X					X	X
II									X	X
III	X	X		X	X	X			X	X
IV	X	X			X	X				X
V	X		X	X			X			X
VI	X		X	X			X		X	X

Table 1. Categorization of the included articles.

The following section summarizes the individual contributions of these articles in terms of the four research questions introduced earlier.

4

Contributions

The contribution of the six articles with respect to the four research questions (RQ column) of this thesis are summarized in Table 2, which is followed by a brief summary of each article's contribution, motivation, goal, methodologies, and results [2].

RQ	No.	Article
1	**I**	Schreiber, R. R., Zylka, M. P. (2020). Social Network Analysis in Software Development Projects: A Systematic Literature Review. In: International Journal of Software Engineering and Knowledge Engineering, Vol. 30, No. 3
2	**II**	Zylka, M. P. & Fischbach, K. (2017). Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda. In: ACM SIGMIS Database, Vol.48, No. 2, pp. 52-78
3	**III**	Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K. & Schoder, D. (2015). Business Model Development, Founders' Social Capital, and the Success of Early-Stage Internet Startups: A Mixed-Method Study. In: Information Systems Journal (ISJ), Vol. 26, No. 5, pp.421-449
3	**IV**	Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Fischbach, K., Schlagwein, D., & Schoder, D. (2014). Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors. In: Proceedings of the Academy of Management Annual Meeting. Philadelphia, PA.
4	**V**	De Oliviera, J. M., Zylka, M. P & Gloor, P. A., Joshi, T. (2019). Mirror, Mirror on the Wall, Who is Leaving of Them All - Predictions for Employee Turnover with Gated Recurrent Neural Networks. In: Studies on Entrepreneurship, Structural Change and Industrial Dynamics. Springer, Cham.
4	**VI**	Gloor, P. A., Zylka, M. P., Fronzetti Colladon, A., Makai, M. (2022). 'Entanglement' – A new dynamic metric to measure team flow. In: Social Networks, Vol. 70, July, pp.100-111

Table 2. Contributions to Research Questions.

[2] Some of the material discussed in this section have, in part, been introduced in the research articles comprising this dissertation, which can be found in the Appendix.

Research Question 1

Article I contributes to the first research question:

What role does social network analysis play in managing projects?

Contribution of Article I

Article I provides a systematic literature review that contributes to Research Question 1 by investigating the current state of research and applications of social network analysis in an organizational context, namely project management. The study was conducted in 2018–19 and was published in 2020 in the *International Journal of Software Engineering and Knowledge Engineering*.

Motivation. There are many different tasks and subtasks involved in the software development process, which can be complex and cover a range of aspects related to software and project management. To help structure and define these tasks and make them more manageable, there are various conceptual models that provide guidance regarding the order in which tasks should be completed. Each approach, however, has its own strengths and limitations and may not be suitable for all types of software development projects. The complexity of a project is an important factor to consider when selecting an appropriate software development model. Organizational network analysis or social network analysis in general can provide insights about the software development process. There has been a significant amount of research on the role of social network analysis in open-source software development projects, but less on the importance and influence of social networks in closed-source software development projects. In contrast to open-source projects, closed-source projects in enterprises may have more flexibility in terms of organizational structure, spatial distribution, task allocation, and competence distribution.

Goal. The goal of this study was to provide a comprehensive overview of how social network analysis is currently being used in software development projects. Additionally, we proposed a research agenda for further exploring the role of social networks in software development projects through a systematic literature review.

Methodology. To identify potential gaps in research and the potential for using social network analysis in software development projects, a systematic literature review was conducted. The systematic literature review is a structured and repeatable method for identifying, evaluating, and interpreting relevant literature on a specific research topic. We followed guidelines for conducting systematic literature reviews in software engineering and focused on research outcomes and methods related to software development projects and social network analysis. A focus on propositions and concepts in these two areas informed the organization of this paper, which is structured conceptually. To present a neutral representation of the review results, no particular perspective was adopted. The intended audience for this review is scholars with a specialized interest in research on using social network analysis in the context of software development projects. The review process was divided into three phases, as suggested by Kitchenham and Charters (2007): planning, conducting, and reporting. These phases and their sub-steps ensure rigor and transparency throughout the review process (see Figure 2).

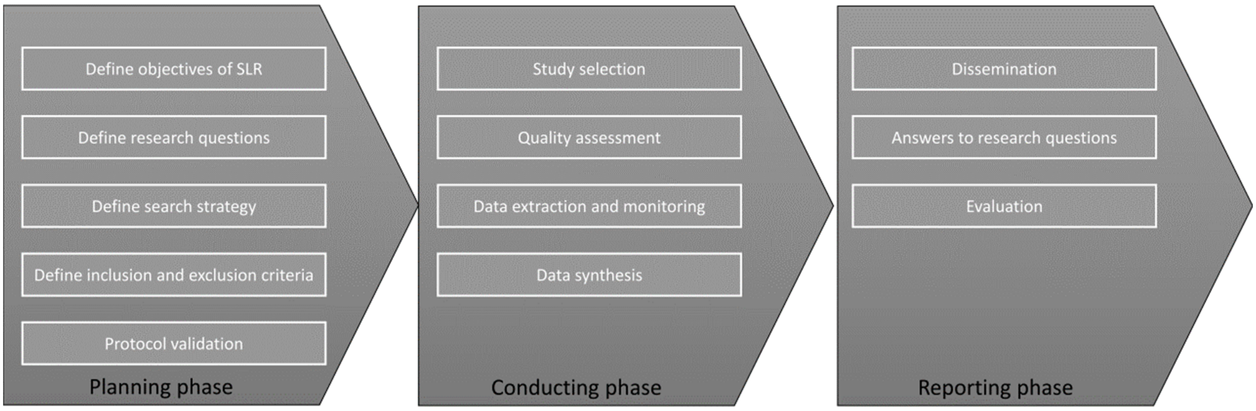


Figure 2: Main literature review phases.

Results. In software development projects, organizing distributed software developers can be challenging due to factors such as geographic distance, cultural differences, different time zones, and inter-company cooperation. From the analysis, we found that research in this area has examined coordination and cooperation in software development projects, particularly in terms of social project structure and clique analysis. Clique analysis includes topics such as core periphery (core team and enhanced team) and clusters. There has also been research on the influence of social capital and power law distribution in software development projects. Previous studies have explored the influence of social capital on the success and developer productivity of software projects and found that cohesive social ties within team members' social networks can lead to increased productivity. In addition, there is evidence that power law distribution exists in software development projects. In this area, many social network analysis studies have focused on the rational characteristics of social ties.

Effective communication within projects is crucial for project success, and it can be complex to manage. Several studies we have found, used social network analysis to visualize and analyze the communication structure between team members. Research on communication networks in large commercial software projects found that software developers perform better when they are central in their team's communication network and when they are embedded in dense communication clusters at the team and project levels. Another study examined how communication develops over time between individuals and groups and its correlation to project development. Using social network analysis to visualize communication patterns can help teams and managers track progress, improve communication, and avoid communication gaps in software development projects.

Further findings from our review have demonstrated the importance of network ties for knowledge flow in software development projects. Research has examined the direction of knowledge flow among different projects in organizational networks and its impact on project success. Network analysis techniques can also be used to identify key authors and subject matter experts, and it has been demonstrated that it is possible to identify knowledge experts in software projects using social network analysis methods. Moreover, boundary spanners who have extensive domain knowledge and hold key positions in the control project structure have been found to be important in this area.

Research Question 2

Article II contributes to the second research question:

What are the potential positive and negative consequences of employee turnover for employees and companies?

Contribution of Article II

Article II is a scoping literature review, which helps to understand the consequences of turnover behavior of information technology employees. This article contributes to the understanding of employee turnover by identifying its antecedents and consequences, as well as highlighting the need for further research into its consequences and identifying research gaps. Through this understanding, top management can effectively weigh the pros and cons of turnover and plan for it as a way to manage human resources more effectively.

Motivation. The demand for highly skilled personnel has been on the rise in recent years due to the increasing importance of technology in various industries. This trend has been fueled by the convergence of the information and communication technology sectors, which has created a unified market that requires a highly skilled information technology workforce. As a result, organizations face intense competition for and the need to retain talented information technology employees. If they are unable to do so, they may suffer from negative consequences such as poor job attitudes, high training and recruitment costs, operational disruption, and a decrease in human and social capital, which can impact performance and profits. Losing a highly skilled employee to a competitor may also bring positive outcomes, such as access to external knowledge for the former employer. With an awareness of the antecedents and consequences of turnover, top management can weigh its pros and cons and effectively manage human resources and might lead to a better more focused applications of organizational network analysis and people analytics.

Goal. The goal of this paper was to provide a taxonomic overview of individual turnover consequences of information technology employees and non-information technology employees that can be utilized in future people analytics research. Further, it aimed to identify research gaps and propose a research agenda for studying these consequences in detail.

Methodology. We conducted a scoping review to examine the current state of research on a particular topic. Scoping reviews are a type of literature review that aim to provide an overview of all available research on a topic, rather than focusing on a specific aspect in depth. They are typically used to identify gaps in the existing research and to inform the design of future studies. The review process followed guidelines outlined by Paré et al. (2016), which emphasize transparency and systematicity. These guidelines provide a framework for conducting literature reviews in a comprehensive and replicable manner. To ensure that our review was as systematic and transparent as possible, we followed a detailed, six-step process and reported on all steps of the review process.

Results. Employee turnover, or the decision to leave a job voluntarily, is a phenomenon that can have both negative and positive impacts. The traditional view is that employee turnover has negative consequences (Dalton & Todor, 1979), both for the individual who leaves and the organization being left. It can be expensive for organizations, as it can lead to increased recruitment and training costs (Dess & Shaw, 2001; Mobley, 1982). Turnover can also disrupt the operations of an organization and can lead to demoralization among individuals (Allen et al., 2001; Staw, 1980). It can have negative impacts on both individuals and organizations (Mobley, 1982; Muchinsky & Morrow, 1980). However, in some cases, turnover can also have positive consequences, not only for the individual who leaves but also for co-workers and organizations (Mobley, 1982). For instance, employee turnover can sometimes be a positive by resolving conflicts and creating promotion opportunities (Dalton & Todor, 1979). It can lead to increased job satisfaction and commitment for the departing employee, and may be viewed as a routine career event by those who have had more frequent voluntary departures. In addition, when new employees join an organization, they may bring with them social capital in the form of relationships with former clients (Somaya et al., 2008). This can facilitate the flow of knowledge between the new and former employer, as the new employee enhances social networks between the two organizations (Corredoira & Rosenkopf, 2010). This can be beneficial to both the new and former employer (Aime et al., 2010). The article shows that current research considers consequences of individual turnover that have not been considered in the specific context of turnover of information technology employees. Despite the longstanding research on the consequences of employee turnover dating back to Price (1977), the specific context of turnover of information technology employees and its consequences has not been adequately considered.

Research Question 3

Articles III and IV contribute to the third research question:

How can social network analysis be applied to assess the consequences of employee turnover for interpersonal and interorganizational networks relevant for people analytics?

Contribution of Article III

The findings of this study provide evidence that external networks are a critical resource for firms at the organizational level, as previously suggested by Canina et al. (2012) and Gnyawali and Madhavan (2001). In particular, the study shows how early-stage startups leverage the information, resources, and status benefits provided by the founders' networks to develop their business models. The study also found empirical evidence that founders' social capital, built from their turnover history, serves as an antecedent to valuable network connections (e.g., to customers and suppliers). This warrants further investigation into the link between individual-level and organizational-level networks in the context of business model development, organizational development, and people analytics. The study is also notable for its use of a combination of qualitative and social network analysis methods that have been found to be mutually informative and provide the same empirical findings. This may provide a useful example for researchers interested in using a combination of these methods in the field of information systems research.

Motivation. There was a lack of empirical research on business models in information systems (IS), as noted by Zott et al. (2011) and Veit et al. (2014), as well as a lack of understanding about the relationship between business models and firm performance and the process of business model change and development (Al-Debei & Avison, 2010; George & Bock, 2011). Prior research had reported that startups often adapt and change their business models (McGrath, 2010; Teece, 2010). Previous research on individual entrepreneurs in various contexts has found that the founders' social capital (e.g., their social networks and personal relationships) can be beneficial in finding new business opportunities. This is particularly relevant for early-stage internet startups, as social capital may provide founders with access to essential knowledge, resources, and investors.

Goal. The aim of this paper was to apply social network analysis to understand the factors that contribute to the success or failure of early-stage internet startups in achieving Series A funding. We hoped to gain a better understanding of the internal processes and characteristics of successful early-stage internet startups.

Methodology. This article utilizes a sequential mixed-method design, which involves using both qualitative and quantitative methods in a specific order. Such a design is used often in information systems research because it allows for a more comprehensive understanding of complex technological, organizational, and social phenomena. Some researchers consider any investigation that utilizes more than one method to be mixed-methods research, while others specifically refer to the combination of qualitative and quantitative methods. This study corresponds to both definitions. We employed a sequential design, which involves first conducting open-ended interviews with experts—here, internet startup founders—to explore qualitatively business model development and the impact of social networks on early-stage internet startups. We developed an interview guideline that was continually updated to reflect new insights or questions. A *purposeful sampling* strategy (Shadish et al., 2001) was employed to select and approach the interviewees, with the goal of including a variety of experiences, locations, and business model types within the specific boundaries of consumer internet startups in or shortly after the early stage.

The study approach followed that outlined by Lichtman (2013) for coding and analyzing the data from the interviews. The initial coding was done by one author, using both *in vivo* codes (terms used by the interviewees) and extant theory (terms from the academic literature). The codes were then merged, and any differences in understanding were resolved through discussion and clarification within the research team. Next, the codes were aggregated into categories, which were higher-level constructs that described several codes at a more abstract level. The categories were refined and organized into main concepts related to the research aim through re-reading of the interviews and identifying relationships between and within categories.

The qualitative findings informed the study's quantitative component. Specifically, our qualitative study enabled us to formulate propositions and turn two of them into a testable hypothesis using natural quantitative data. Social network analysis was then used to test the hypothesis that social networks of founders have a positive effect on the success of their startups. We used eigenvector centrality as a measure of social capital. This measure, recommended by Borgatti et al. (1998), is widely used in social network analysis. We opted for the directed unweighted scaled eigenvector centrality, as opposed to the directed weighted version, in order to take into account all previous employers and increase the diversity of inbound flows rather than focusing solely on high-intensity turnover relationships. This allowed for calculating the directed unweighted scaled eigenvector centrality for each startup in the network. The findings of this analysis of the quantitative data provided further evidence to support the hypothesis that the combined social capital of the founders (as measured by their startup's eigenvector centrality) has a significant positive impact on the success of their early-stage internet startup (as measured by obtaining Series A funding).

Results. We carried out a t-test to determine if there is a statistically meaningful difference between the group of successful startups and the group of unsuccessful startups (see Table 3). The result of the t-test confirmed that the means of the two groups are significantly different. Thus, the hypothesis was supported by the data at the 0.01 significance level. In addition to testing the hypothesis (i.e., the impact of founders' social networks), we also controlled for some other factors that could potentially affect funding success. These include typical human capital factors such as education and work experience (Lazear, 2004), as well as specific entrepreneurial (C. Chen et al., 1998) and technological skills (Compeau & Higgins, 1995). We found that founders' experience and specific skills did not have a statistically meaningful impact on the success of a startup. The only exception to this was the number of years of education, which showed a moderate level of significance.

Category	**Measure**	**Successful startups**	**Unsuccessful startups**	**Significance of difference**	**Test type**
Social capital	Founders' network centrality	21 (30%)	49 (70%)	0.002***	b
Human capital factors	Average number of (co-) Founders	2.238	2.000	0.270	b
	VC experience	3 (14%)	4 (8%)	0.421	a
	Consultant experience	6 (28%)	17 (34%)	0.783	a
	Founder experience	10 (47%)	34 (69%)	0.108	a
	Average years of education	4.786	3.699	0.074*	b
	Average years of work experience	9.572	9.446	0.936	b

Entrepreneurial self-efficacy	Management	19 (90%)	40 (81%)	0.485	a
	Financial	15 (71%)	36 (73%)	1.000	a
	Risk	16 (76%)	38 (77%)	1.000	a
	Innovation	11 (52%)	30 (61%)	0.599	a
	Marketing	13 (61%)	29 (59%)	1.000	a
Computer efficacy	General IT	15 (71%)	31 (63%)	0.590	a
	Web 2.0	9 (42%)	24 (48%)	0.795	a
	Design	2 (9%)	10 (20%)	0.325	a
a = Fisher's exact test; b = *t*-test (2-sided); significant at * α = 0.100; ** α =0.050, *** α =0.010					

Table 3: Results from tests of association (N=70, showing p-values and significance levels).

Contribution of Article IV

Article IV was the result of an empirical study that exploited data on employee turnover in an explorative approach to derive an industry classification system. The paper was presented at the Annual Meeting of the Academy of Management in 2014. It contributes to the third research question by exploring the capabilities of social network analysis, resulting in a novel approach to industry classification. It provides valuable insights into the structure of network formed by the organizations analyzed and their turnover relationships, demonstrating the potential for using social network analysis on a macro-level network. Further, it contributes to people analytics. We showed that there are different patterns of employee turnover depending on the sector. Thus, considering where employees potentially come from and are likely to go to will provide more insights and allow, for example, for more targeted measures to prevent loss of skilled employees. Further, the methodology presented in the article could be used to assess the effects of employee turnover on the organizational performance from a network perspective.

Motivation. Chiasson and Davidson (2005) conducted a review of research articles published in the journals *MIS Quarterly* and *Information Systems Research* between 1997 and 2004 and found that a majority (58%) of the studies they analyzed did not specify the industry in which the study was conducted. They pointed out that industry context is important for the development of new information systems theory and for understanding the limitations of existing information systems theory. A later review by

Seddon and Scheepers (2011) of articles published in *MIS Quarterly* and *Information Systems Research* in 2007 and 2008 found that only 24% of the empirical papers included discussion of the boundary conditions that were relevant to their results.

Failing to consider the industry context in a study can have a significant impact on the research results by inadequately defining the theory or by resulting in generalizations that may not be relevant in other settings (Chiasson & Davidson, 2005). Chiasson and Davidson (2005) also noted that identifying the specific industry or industries relevant to a study is both an empirical and conceptual issue. While broad industry categories such as "manufacturing" and "service sector" may be useful for meta-analyses, more detailed classification schemes such as the North American Industry Classification System can be helpful in defining industry boundaries (Xue et al., 2011). It is important to bear in mind, however, that classification systems such as the International Standard Industrial Classification and North American Industry Classification System, which are commonly used in statistical tests, are only as valid as the underlying classification system. These systems may not always accurately reflect the complex and diverse nature of industries, particularly given the increasing convergence of technology and industries (Garcia-Murillo & MacInnes, 2003) and the widespread use of information and communication technology across various sectors (Karmarkar & Apte, 2007).

Goal. The purpose of this paper is to investigate a classification method by integrating employee turnover data and social network analysis to address some of the problems mentioned above.

Methodology. Industries can be studied from three distinct yet complementary research perspectives: economic, institutional, and social/cultural (Crowston & Myers, 2004). The economic perspective is the most widely used approach in information systems research, however, the other two perspectives have several advantages. The social/cultural perspective highlights the social connections, networks, and frameworks that can be created by the way people and organizations interact within a particular industry (Crowston & Myers, 2004). This includes employee turnover, which has been identified as an area of interest from a social network analysis perspective; Collet and Hedström (2013) found that connections in inter-organizational networks established through employee turnover is influenced by the direction of previous connections and that the majority of knowledge exchange that results from this movement happens over short sociometric distances. Employee turnover can also be used to identify the existence and function of an industrial cluster.

There have been few systematic efforts to define industrial systems based on turnover flows, with the exception of some case-based economic geographical studies (Agrawal et al., 2003). In this study, data on employee turnover were collected from the social media platform LinkedIn, which provides such data in the form of employment histories submitted by its members. At the time of the study, LinkedIn offered this information in the form of aggregated flows of employees between organizations. A snowball sampling method was used to identify information systems-related organizations and the employee turnover flows between 29,419 organizations were collected in June 2012. These data were used to create a network in which organizations were represented as nodes and the directed edges between them represented turnover of employees, with the number of employees moving from one organization to another serving as the weight. To identify industry clusters within this network structure, the community detection algorithm developed by Rosvall and Bergstrom (2008) was used. This algorithm is particularly well suited for large networks with weighted and directed ties and has demonstrated good performance in several benchmark studies (Lancichinetti & Fortunato, 2009). By applying the algorithm to the turnover network, it was possible to identify a hierarchy of clusters based on employee turnover flows. As the algorithm aims to maximize the internal cohesion of these clusters, it was expected that they would accurately reflect the dynamics of employee movements, as described by Collet and Hedström (2013). The resulting industry clusters at the two highest levels of the hierarchy were manually reviewed and labeled based on their composition.

Results. In total, the algorithm identified 135 clusters at the root level. The distribution was heavily skewed, with the first 10 clusters accounting for 97.42% of the companies. The top-3 clusters displayed distinct patterns in their underlying classification. Cluster 1 was primarily composed of companies in information and communication technology-related classifications, with some companies from the internet sector. However, there were almost no companies from the content and media sector. In contrast, Clusters 2 and 3 were almost exclusively composed of companies from the content and media sector.

From an analysis of the characteristics of the information economy sectors based on Organization for Economic Co-operation and Development definitions and the International Standard Industrial Classification system, it was found that there is significantly more turnover within the information economy compared to industries outside of it. In addition, the flow of employees in and out of the information economy is almost balanced. When examining the industries and companies that exchange employees with the information economy, it was observed that there are significant differences between the three sectors. For example, an analysis of the top-10 sources of employees for each sector revealed some interesting patterns: the information and communication technology sector receives employees from outsourcing partners and the military; the content and media sector receives employees primarily from financial companies; and the Internet sector receives employees from universities and consulting firms. These differences between the sectors are also apparent when considering company size, type, and age. The internet sector, for instance, is characterized by smaller and younger, predominantly privately held companies, while the information and communication technology and content and media sectors have many more companies with more than 10,000 employees and significantly more public companies, reflecting the different levels of maturity in these sectors.

Upon closer examination of the information economy sectors, it was found that the industry is not as homogeneous as one might expect, at least in terms of employee turnover. Specifically, it was observed that there was a greater rate of employee turnover within each industry sector than between sector, which resulted in distinct industry clusters that are also visible when analyzing the network graph. The internet sector has a strong association with the information and communication technology sector; however, it can be clearly distinguished, meanwhile the content and media sector have more close connections to the other two sectors. These clusters are not only present at the sector level but also at the sub-sector level. For instance, within the information and communication technology sector, there are distinct clusters for network equipment, semiconductors, services (consulting and outsourcing), and software and platforms.

Research Question 4

The articles in this section contribute to the fourth research question:

How can social network analysis be applied to predict company-relevant employee behavior patterns for people analytics?

Contribution of Article V

This study provides a novel approach for analyzing longitudinal employee turnover data. It is among the first to use a recurrent neural network beyond the usual applications, such as objects or speech recognition. As such, this paper may provide an insightful account for researchers interested in the context of predictive people analytics and provide an example of how social network analysis measures and a deep learning approach can be meaningfully integrated in people analytics. Further, we addressed the calls by

Hom et al. (2017) and Lee et al. (2017) by considering the dynamic nature of antecedents of employee turnover and conducting a social network analysis, whereas earlier research had primarily used a standard research approach (Steel, 2002). This paper contributes to people analytics by providing a novel analytical perspective (social network analysis + deep learning) on key elements of turnover models. Our experiment offers a way to predict employee turnover, which can be disruptive and detrimental to company performance. By implementing our model into a human resource information system as an *early warning system*, managers could be alerted about the probability of an important and highly skilled employee departing the company, allowing them to implement retention strategies.

Motivation. Employees can significantly impact a firm's productivity and growth through their contribution of new knowledge (Aime et al., 2010; Somaya et al., 2008). Therefore, organizations often invest in human resources in order to benefit from these contributions. If highly skilled employees leave for other opportunities, however, it can diminish these benefits and even lead to negative consequences, such as knowledge leakage to competitors, reduced human capital, and a decline in company performance. To prevent these outcomes, organizations compete to attract and retain highly skilled personnel, a competition referred to as the "war for talent" (Chambers et al., 1998). Consequently, the turnover of highly skilled employees is a topic of interest for practitioners and scholars alike.

The topic of employee turnover, particularly among highly skilled workers, has garnered significant attention from both practitioners and scholars in fields such as applied psychology, human resources, and management. Many studies have focused on understanding the reasons behind actual turnover behavior, with job dissatisfaction being identified as a major proximal cause. Researchers have typically used survey data and statistical analysis techniques such as ordinary least squares regression, survival and hazard functions, and structural equation models to predict turnover. However, the accuracy of these predictions has been found to be only moderate at best, with slightly better results when using turnover intention as a proxy for actual turnover behavior. There is a need for new approaches that consider the evolution of turnover determinants over time, as changes in these factors can affect turnover through their impact on proximal antecedents.

Goal. We followed the call for novel approaches for employee turnover prediction applying a gated recurrent neural network model (Chung et al., 2014), a variation of long short-term memory networks

(Gers, 2001; Hochreiter & Schmidhuber, 1997) on longitudinal turnover data. Our goal was to predict turnover of highly skilled employees.

Methodology. This study analyzed e-mail communication data from a global professional services firm with over 70,000 employees in over 20 countries. The data set included 845,208 actors, including employees and external stakeholders, and approximately 138 million edges representing the communication that occurred from January 1, 2017, to January 29, 2018.

We analyzed e-mail communication data from managerial employees, which comprised 35% of the total e-mail communication data provided. We used network metrics, calculated for 15-day time frames with a 3-day gap between each frame, as input variables for deep learning models. Each comma-separated values file, representing a combination of metrics from 3 months of e-mail communication, was used as a single data point. Employee turnover data were used to classify each data point as either representing an employee who left the company or one who did not. The focus was on the 5-7 months prior to turnover, when we believe employees may be considering leaving a company and may exhibit changes in communication behaviors.

We developed a recurrent neural network with gated recurrent units; it was applied to predict employee turnover at a company. The model was trained on a dataset of 2898 employees and tested on a separate dataset of 954 employees. To ensure the robustness of our model's performance, we applied fourfold cross-validation, dividing the training data into four equal-sized subsets and using one of the subsets as validation data for testing while using the remaining three subsets for training. The dataset was imbalanced, with a small number of employees who had left the company (78 in total). To address this issue, we used the synthetic minority over-sampling technique to oversample the minority class (the employees who had left). Synthetic minority over-sampling technique creates synthetic instances by joining the k-nearest neighbors of a minority class instance, resulting in increased sensitivity to the minority class without excessively oversampling it. To evaluate the model's performance, we used a variety of metrics, including precision, recall, accuracy, the area under the curve, and the Matthews correlation coefficient score, a discretization of the Pearson correlation value.

	Training	Test
# of actors	2,898	954
# of leavers	59	19
# of stayers' time series	238,438	78,468
# of leavers' time series	832	262
# of stayers' synthetic minority over-sampling technique time series	238,438	78,468
# of leavers' synthetic minority over-sampling technique time series	238,438	78,468

Table 4: Statistics of the dataset

Results. We conducted the gated recurrent unit experiment with a keep probability of 0.1, one LSTM layer, eight neurons, and a learning rate of 0.001. All folds performed well; the average fold has an ACC of 0.933 and a Matthews correlation coefficient of 0.873. The test set's performance was good. The accuracy as 0.800, and Matthews correlation coefficient (0.615) showed a strong positive predictive power. The results of the experiments revealed that applying a deep learning approach has potential to conduct a binary classification of employees in stayers or leavers by analyzing their e-mail communication behavior.

	Parameters	Accuracy	Precision	Recall	Area under the curve	Matthews correlation coefficient
Model 1	Keep Probability: 0.1 Layers: 1 Size: 8 Learning Rate: 0.001	0.768	0.653	0.849	0.786	0.554
Model 2	Keep Probability: 0.1 Layers: 1 Size: 16 Learning Rate: 0.001	0.800	0.745	0.844	0.816	0.615

Table 5: Model results based on different configuration (test set, 1.5-fold).

Contribution of Article VI

Article VI contributes to synchronization and flow state research by providing a novel metric for determining communication synchronization in organizational contexts. We validated this metric through four case studies in different organizations. People analytics can use our metrics to measure team flow not only by conducting surveys, but also by using social network analysis and taking the relational perspective into account.

We further contribute to people analytics research with this paper by providing a novel metric for analyzing employee communication and behaviors. Employee communication is a critical factor for good collaboration and employee and team performance (Gloor et al., 2020; Wen et al., 2019), and thus using a new metric of communication dynamics might open new research opportunities. Decisionmakers such as human resources managers could act in an interventionist manner (Valente, 2012), however, and use this metric to identify weak and strong entangled actors in the communication network of a team or in the entire company. Thus, human resource managers might use this metric to improve performance appraisal systems, anticipate disengagement, and improve hiring and retention strategies. Combining novel metrics of e-mail communication analysis with long-established methods to assess employees' satisfaction (like surveys), managers can offer improved organizational initiatives, such as mentoring programs or cross-staffing, or retention strategies.

The entanglement metric introduced in this article has the potential to help managers better understand the nature of employee online communication at their particular organizations. This might lead to a rethinking of team design and team building in the specific organization, which could ultimately lead to improved communication and collaboration and might support the identification of cohesive groups.

Motivation. Team synchronization, or being "in sync" with one's colleagues in a work environment, is a topic of research worth considering for decisionmakers. This is because being in sync can have a positive impact on employee performance, increasing cooperation among team members by strengthening social attachment.

Both offline communication, such as face-to-face meetings, and online communication, such as emails and chat tools, play a role in determining whether a team is in sync or "out of sync." While offline communication is generally more effective at establishing team synchronization, online communication can be hindered by asynchronous characteristics, such as time lags, which may prevent the development of a shared team rhythm. It is possible, though, to analyze online communication data in near-real time to monitor team learning and performance continuously. Metrics based on communication flow and amount of communication are suitable for this purpose.

In addition, research has shown that analyzing online communication data in organizational contexts can be used as a predictor for job-related constructs such as employee turnover and employee performance.

For example, the speed at which an email is responded to has been found to be a good predictor of individual and team performance. Based on these findings, we hypothesized that patterns of team online communication gathered through a social network analysis approach can also be used to show team synchronization. Therefore, we conceptualized a new measure.

Goal. We introduced a metric called entanglement that measures the synchronization of e-mail communication behaviors of team members and their flow state over time. This metric is grounded in social network analysis and identifies the similarity of timeseries of social network analysis metrics. We validated the metric by conducting four case studies, using different datasets from different organizations.

Methodology. To demonstrate the concept and calculation of entanglement using an example, we can consider an individual mailbox representing a dataset of e-mails exchanged between individuals who work on multiple projects together. First, we collected the mailbox data and stored it in a database, organizing the e-mail data in a way that allowed us to view it from a network perspective. From this structured data, we could then analyze the degree of entanglement between the individuals by looking at the patterns of e-mail communication and the ways in which the e-mails linked different individuals and projects together. To determine the degree of entanglement between the mailbox owner and their colleagues, we could analyze the communication activity between them by considering the time series of messages exchanged. Specifically, we could calculate the inverse of the Euclidean distance between the activity time series of each actor in the network. This value will be larger when the activity time series of two actors are more similar.

It is important to consider the context in which actors are communicating in the network. For example, two pairs of actors may exchange the same number of messages, but the first pair may be more tightly connected to other actors in the network, while the second pair has weaker connections to other actors. In this case, the entanglement between the actors in the first pair may be stronger due to their higher level of communication with others in the network. To make this metric comparable among pairs of actors with different levels of activity in the same network, we multiplied it by the product of the degree centralities of both actors. Degree measures the centrality, sometimes seen as a proxy of popularity, of a node in a network, by counting the number of its nearest neighbors (Freeman, 1977). Further it can be a proxy for the level of engagement within a group, team or organization (Gloor et al., 2020).

Communication activity via e-mail (Gloor et al., 2014) indicates the number of e-mail messages sent by a person within a time interval.

Figure 3 shows the e-mail communication activity over a period of time for the email box we analyzed. The blue line represents the communication activity of the mailbox owner; the other lines represent the communication activity of the individuals with whom the owner exchanged the most emails. When the communication activity between the owner and another person is highly correlated, it means they are in sync and share a similar flow of activity over time, which indicates a strong level of entanglement between them. The graph demonstrates the importance of considering degree centrality when calculating entanglement. This is because levels of communication activity can vary significantly between different actors, even if they are exchanging messages at the same time.

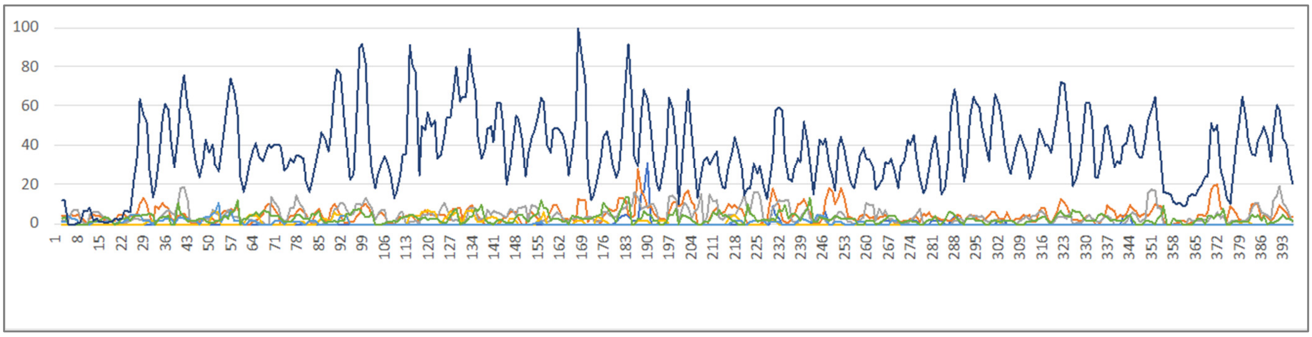


Figure 3: Flow of e-mail communication by time.

Accordingly, we defined the activity entanglement $E_A(x_T, y_T)$ between two individuals, named x and y in a specific time window T, as:

$$E_A(x_T, y_T) = \frac{C_D(x_T)\, C_D(y_T)}{d\big(A(x_T), A(y_T)\big)}$$

where $C_D(x_T)$ and $C_D(y_T)$ are the degree centralities of the two individuals x and y, and $d\big(A(x_T), A(y_T)\big)$ is their Euclidean distance, with respect to communication activity A in a defined time window T. In other words, the entanglement of two individuals x and y is given by the multiplication of the number of their direct contacts in the e-mail network divided by their synchronization of communication activity.

As indicated above, it is necessary to include the product of the degree centralities of x and y into the entanglement formula to provide for the differences in centralities between actors: assume that actor x has low degree, if x is synchronized with highly connected actor y having high degree centrality, the high degree of actor y will boost entanglement of actor x in comparison with all other actors in the network. In other words, we wanted our metric to reward less-influential actors who are synchronized with influential actors. In addition to considering communication activity, we could also consider synchronization in weighted and unweighted betweenness centrality as a factor in determining entanglement.

After the conceptualization of the new metric and its variants, we conducted four case studies, with different datasets from different organizations, to validate the metric. In each case, the same data collection process was followed. This involved collecting the emails of a sample of project members who agreed to participate in the pilot studies. All of these individuals were employed at large organizations at the time of the data collection. The email data was normalized for time zones, and the time window for calculations was set to 7 days, as this has been found to be most effective for this type of organizational email data. The relationship between entanglement, as calculated from email communication, and individual and group outcome variables were measured. The focus of the analysis was on communication-based measures, such as the number of messages sent and received, and network centrality measures, as the goal was to explore the properties of communication networks. In addition, we used online communication metrics that were specifically designed to evaluate interactivity in email communication. Specifically, we examined communication activity, which reflects the number of emails sent by a person within a specific time period, and the number of nudges, which represents the average number of emails that a sender needs to send in order to receive a response from the receiver. We distinguished between ego nudges (the number of emails required for the recipient to respond) and alter nudges (the number of emails required for others to respond). Further, we measured the contribution index, which is the balance between messages sent and messages received (Gloor, 2017). Finally, we calculated the average response times to measure the length of time it took a person to reply to an email. This metric is useful for identifying fast and slow communicators and for identifying patterns of behavior by looking at periods of slower response. We calculated both ego average response time, which represents the average number of hours it takes a sender to respond to emails, and alter average response time, which reflects the average number of hours it takes recipients to respond to a sender.

We presented four case studies to demonstrate how the proposed entanglement metric can be used with email data to predict work-related outcome variables, such as team performance and employee turnover (see Table 6). These case studies are related to different business contexts and consider different dependent variables. In all cases, we analyzed email data to demonstrate the suitability of the entanglement metric for online communication data. Our objective was not to compare results directly across the case studies, draw general conclusions, or claim causality, but rather to illustrate the versatility of our entanglement metric, which can be applied to study business interaction dynamics in a variety of scenarios.

Case study	Industry	Research object	Entanglement measure	Entanglement level	Outcome variable
A	Health care	53 employees in 11 healthcare innovation teams	Activity entanglement	Team	Team performance & learning behavior
B	Professional services	113 senior executives	Activity entanglement	Individual	Employee turnover
C	Professional services	81 managers	Betweenness entanglement	Individual	Employee performance
D	Professional services	82 mangers in 13 teams	Group betweenness entanglement	Team	Customer satisfaction

Table 6. Applied case studies in Article VI.

Results. We demonstrated through four case studies, using real-world datasets, that our proposed metric and its variations are good predictors of various individual and team performance indicators (a summary of our results can be found in Table 7). First, we discovered that the dispersion of activity entanglement is positively correlated with team performance. This means that the synchronized communication activity of certain team members and their sustained similar flow state can enhance the performance of the team. These findings are similar to studies showing that email communication and face-to-face communication frequency (Patrashkova-Volzdoska et al., 2003), as well as flow in knowledge work (Quinn, 2005), can all contribute to higher team performance. It also seemed that the best teams exhibit higher dispersion, comprising highly entangled team members and more peripheral ones. Teams might benefit from strong leadership of few selected individuals that can guide and inspire others. Other research has demonstrated that communication-based measures of social network analysis can be used to predict voluntary employee turnover (see Article V, Gloor et al., 2017). Our own research has shown that the metric of entanglement can also accurately predict employee turnover and may improve the accuracy of these types of models. In terms of employee disengagement, this metric could potentially be useful in understanding and addressing the issue.

Case study	Dependent variable	Result summary
A	Team performance and learning behavior	The wider the spread in activity entanglement of the team members, the higher the team performance and learning behavior. This corresponds to having some core team members strongly entangled and the remaining members weakly entangled.
B	Employee turnover	The Gini index of activity entanglement is the variable with the highest impact on model predictions. Employees who stay in the company have high Gini entanglement probably using selective communication and interacting more with some colleagues than with all others. They are also more responsive to emails and take less time to answer.
C	Individual performance	Tenure, betweenness centrality, and Gini entanglement are the most important predictors of top performers. – with high Gini index of betweenness entanglement and high betweenness centrality significantly increasing the chance of being classified as a top performer.
D	Customer satisfaction	The Gini index of group betweenness entanglement for teams is related to customer satisfaction. The higher the inequality in group betweenness entanglement is for a team, the happier its customer. This means that customers are happier when a few entangled leaders emerge in the team.

Table 7: Results of Article VI.

Second, we also discovered that the Gini coefficient of betweenness entanglement and betweenness centrality are related to individual employee performance. A high Gini index of betweenness entanglement significantly increases the likelihood of being a top performer, indicating that focused communication—intensive and highly synchronized communication with a select group of colleagues while reducing communication with the rest of the organization—is a sign of high performance. These findings align with previous research (Brass, 1984; Mehra et al., 2001; Sparrowe et al., 2001) which has shown that network centrality is positively related to individual performance. However, an important aspect of our metric is that synchronization with others has a positive impact on individual performance, not just having a central social position. Centrality alone may not fully explain individual performance (Reinholt et al., 2011), and our betweenness entanglement metric addresses this issue. Additionally, we found that low tenure also has a positive effect on individual performance.

Third, we found that inequality of group betweenness entanglement within teams has a positive impact on customer satisfaction. The company in case study D uses customer satisfaction/net promoter score as a measure of team performance. Our findings suggest that teams with stronger leaders who have high entanglement tend to have happier customers, indicating that these leaders have a significant influence on team dynamics over time, while the rest of the team may be more passive. Previous research by Mukherjee (2016) has shown that centralized leadership can have a positive effect on sport team performance, but Mehra et al. (2006) suggests that distributed leadership structures can differ in important structural characteristics and these differences can either positively or negatively impact team performance (Cummings & Cross, 2003).

5

Limitations

This section provides a general overview of the overarching limitations of the research discussed in this dissertation, but does not summarize the limitations of each article—for which readers are referred to each article's own limitations section.

This thesis addresses an important research gap and reveals several areas that require further research. While this work provides an initial conceptual foundation at the intersection between social network analysis and people analytics, it points toward the need for a more thorough investigation of the two fields. This would improve future research at that intersection and allow for a more systematic and comprehensive transfer of theories and methods of social network analysis to the people analytics domain. As mentioned in section 3, people analytics could benefit from network theory and a greater understanding of the mechanisms and processes that interact with network structures to yield certain outcomes for employees and teams, which can generate insights useful in making strategic decisions and improving organizational effectiveness, efficiency, and outcomes.

Further, the classification of the articles shows that there is still a need for prescriptive people analytics studies. This aligns with the findings of Giermindl et al. (2022), who found that most people analytics studies examine only the use of descriptive and predictive analytics, but "recent technological advances in the context of people analytics illustrate the need to expand current research and emphasize the use of more advanced forms of people analytics" (Giermindl et al., 2022, p. 417).

While this thesis focuses on exploring the positive potential of social network analysis in people analytics, it also recognizes the importance of addressing ethical considerations in the use of this approach. Future research could benefit from a more thorough analysis of the ethical implications of using social network analysis in people analytics (Gal et al., 2020; Tursunbayeva et al., 2022).

6

Conclusions

This thesis contributes to understanding the role social network analysis could play in people analytics. I presented an overview of the theoretical background of social network analysis and its application in organizational contexts. Further, I linked people analytics with social network analysis through the relational lens of network theory and showed how the predictive power of people analytics can be improved. Building on this, I have provided a framework that can be used to classify organizational studies that apply social network analysis for people analytics purposes. I contrasted the characteristics of such studies with traditional investigations of social networks and derived four research questions from the findings:

RQ1	What role does social network analysis play in managing projects?
RQ2	What are the potential positive and negative consequences of employee turnover for employees and companies?
RQ3	How can social network analysis be applied to assess the consequences of employee turnover for interpersonal and interorganizational networks relevant for people analytics?
RQ4	How can social network analysis be applied to predict company-relevant employee behavior patterns for people analytics?

Table 8: Main research questions

As a group of six, the papers presented in this dissertation offer contributions to all four research questions. Two of the articles (Article I & II) are literature reviews. Article I focused on social network analysis in project management, whereas Article II investigated employee turnover and its consequences on a micro and macro level. Articles III and IV provide new applications of state-based data (employee turnover data) in organizational contexts. Articles V and VI contribute to predictive analyses of employee turnover with social network analysis.

With respect to the first research question, Article I concludes that social network analysis provides a crucial set of methods for analyzing the structure, position, autonomy, and relationship of individuals in software development projects. In addition, the topics of project organization, communication management, knowledge management, version and configuration management, requirement management, and risk management provide a wide range of research opportunities.

As for the second research question, Article II examined the current research state of individual turnover behavior consequences. Building upon the existing turnover literature from different disciplines, the article revealed that there is a relatively vague understanding of the consequences of individual voluntary turnover of information technology employees. The findings, especially the research gaps we identified, have enabled us to establish a research agenda for turnover behavior research with a focus on the consequences. In addition, our findings call attention to problems human resources management must address and resolve. Effective management of human resources demands awareness of the antecedents and the consequences of turnover.

With respect to the third research question, Article III found strong support for the idea that startups with better-connected founders are more successful because the founders' professional social networks provide the required means (i.e., information, resource, and status benefits) to develop the business model in the early stage of the startup. We showed this link based on 17 expert interviews with founders around the globe and through the application of social network analysis to a unique sample of 70 startups and the professional networks of their 145 founders. More research needs to be done on business models, but the findings we presented contribute significantly to understanding how business models evolve and affect the success of early-stage internet startups. In Article IV, we examined employee turnover and mobility within the information economy industry, with a focus on three subsectors. Our findings indicate that these subsectors have distinct characteristics related to employee mobility and company size, type, and age, suggesting that the information economy industry is not uniform from a turnover perspective. This article has helped in developing a method for classifying sectors and sub-sectors of the information economy using real-world data, specifically employee turnover data.

As for the fourth research question, we applied a gated recurrent unit classifier that classifies employees in two states (leaver or stayer) by taking their e-mail communication data into account (Article V). The classifier's performance was evaluated using a variety of metrics, including accuracy, recall, precision, the area under the curve, and Matthews correlation coefficient values. The recurrent neural network model we developed provides promising performance. Here, a gated recurrent unit can strongly benefit from the fact that it can look back in time and learn to correlate the calculated network metrics. Gated recurrent unit approaches can learn these correlations, although it might require further training or different variables to be added to the data. We concluded that our trained gated recurrent unit classifier is very suitable for predicting employees' turnover behavior. This is the first reported demonstration of a successful application of gated recurrent unit neural networks to data from an organizational management context. Article VI showed that the idea of synchronization and flow state can be used together to develop new metrics—based on methods and tools of social network analysis. The findings from our four case studies give evidence to the potential of our proposed entanglement metric. We positioned our research as a starting point for further people analytics research, which should consider employees' social interactions and communication, with the goal to improve and optimize collaboration, leading to more satisfied employees and customers.

7

References

Agrawal, A., Cockburn, I. M., & McHale, J. (2003). Gone but not forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital. In *NBER Working Paper Series* (No. 9950; NBER Working Paper Series, Issue September). http://www.nber.org/papers/w9950

Aime, F., Johnson, S., Ridge, J. W., & Hill, A. D. (2010). The routine may be stable but the advantage is not: competitive implications of key employee mobility. *Strategic Management Journal*, *31*(1), 75–87. https://doi.org/10.1002/smj.809

Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, *19*(3), 359–376. https://doi.org/10.1057/ejis.2010.21

Allen, T. D., Freeman, D. M., Russell, J. E. A., Reizenstein, R. C., & Rentz, J. O. (2001). Survivor reactions to organizational downsizing: Does time ease the pain? *Journal of Occupational and Organizational Psychology*, *74*(2), 145–164. https://doi.org/10.1348/096317901167299

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science*, *58*(5), 913–931. https://doi.org/10.1287/mnsc.1110.1460

Argyris, C. (1960). Understanding organizational behavior. Dorsey.

Baker, W. E. (1992). The network organization in theory and practice. In N. Nohria & R. Eccles (Eds.), *Networks and Organizations* (pp. 327–429). Harvard Business School Press.

Barabási, A. (2002). *Linked: The new science of networks*. Perseus Books Group.

Borgatti, S. P., Candance, J., & Everett, M. G. (1998). Network Measures of Social Capital. *Connections*, *21*, 27–36.

Borgatti, S. P., & Foster, P. C. (2003). The Network Paradigm in Organizational Research: A Review and Typology. *Journal of Management*, *29*(6), 991–1013. https://doi.org/10.1016/S0149-2063(03)00087-4

Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. *Organization Science*, *22*(5), 1168–1181. https://doi.org/10.1287/orsc.1100.0641

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(February), 892–896.

Borgo, D. (2006). Sync or Swarm: Musical Improvisation and the Complex Dynamics of Group Creativity. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 4060 LNCS (pp. 1–24). https://doi.org/10.1007/11780274_1

Boudreau, J., & Cascio, W. (2017). Human capital analytics: why are we not there? *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 119–126. https://doi.org/10.1108/JOEPP-03-2017-0021

Brass, D. J. (1981). Structural Relationships, Job Characteristics, and Worker Satisfaction and Performance. *Administrative Science Quarterly*, *26*(3), 331. https://doi.org/10.2307/2392511

Brass, D. J. (1984). Being in the Right Place: A Structural Analysis of Individual Influence in an Organization. *Administrative Science Quarterly*, *29*(4), 518. https://doi.org/10.2307/2392937

Brass, D. J. (2003). A social network perspective on human resources management. In R. Cross, A. Parker, & L. Sasson (Eds.), *Networks in the knowledge economy* (pp. 283–323). Oxford University Press.

Brass, D. J. (2022). New Developments in Social Network Analysis. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*(1), 225–246. https://doi.org/10.1146/annurev-orgpsych-012420-090628

Burt, R. S. (1992). Structural Holes: The Social Structure of Competition. Harvard University Press.

Burt, R. S. (1997). The Contingent Value of Social Capital. *Administrative Science Quarterly*, *42*(2), 339. https://doi.org/10.2307/2393923

Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, *325*(July), 414–416. https://doi.org/10.1126/science.1171022

Canina, L., Palacios, D., & Devece, C. (2012). Management theories linking individual and organizational level analysis in entrepreneurship research. *International Entrepreneurship and Management Journal*, *8*(3), 271–284. https://doi.org/10.1007/s11365-010-0166-8

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis*. Cambridge University Press.

Centola, D., & Macy, M. (2007). Complex Contagions and the Weakness of Long Ties. *American Journal of Sociology*, *113*(3), 702–734. https://doi.org/10.1086/521848

Chambers, E. G., Foulon, M., Handfield-Jones, H., Hankin, S. M., Iii, E. G. M., & Michaels III, E. G. (1998). War for Talent. *The McKinsey Quarterly*, *3*, 44–57. https://www.mckinseyquarterly.com/The_war_for_talent_305

Chamorro-Premuzic, T., & Bailie, I. (2020). Tech Is Transforming People Analytics. Is That a Good Thing? *Harvard Business Review*. https://hbr.org/2020/10/tech-is-transforming-people-analytics-is-that-a-good-thing

Chellappa, R. K., & Saraf, N. (2010). Alliances, Rivalry, and Firm Performance in Enterprise Systems Software Markets: A Social Network Approach. *Information Systems Research*, *21*(4), 849–871. https://doi.org/10.1287/isre.1090.0278

Chen, C., Greene, P. G., & Crick, A. (1998). Does entrepreneurial self-efficacy distinguish entrepreneurs from managers? *Journal of Business Venturing*, *13*(4), 295–316. https://doi.org/10.1016/S0883-9026(97)00029-3

Chen, H., Mehra, A., Tasselli, S., & Borgatti, S. P. (2022). Network Dynamics and Organizations: A Review and Research Agenda. *Journal of Management*, *48*(6), 1602–1660. https://doi.org/10.1177/01492063211063218

Chiasson, M. W., & Davidson, E. J. (2005). Taking Industry Seriously in Information Systems Research. *MIS Quarterly*, *29*(4), 591–605.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. http://arxiv.org/abs/1412.3555

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*(1988), 95–120.

Collet, F., & Hedström, P. (2013). Old friends and new acquaintances: Tie formation mechanisms in an interorganizational network generated by employee mobility. *Social Networks*, *35*(3), 288–299. https://doi.org/10.1016/j.socnet.2013.02.005

Compeau, D. R., & Higgins, C. A. (1995). Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly*, *19*(2), 189. https://doi.org/10.2307/249688

Corredoira, R. A., & Rosenkopf, L. (2010). Should auld acquaintance be forgot? The reverse transfer of knowledge through mobility ties. *Strategic Management Journal*, *31*(2), 159–181. https://doi.org/10.1002/smj.803

Cross, R., & Parker, A. (2004). The hidden power of social networks: Understanding how work really gets done in organizations. Harvard Business Press.

Crowston, K., & Myers, M. D. (2004). Information technology and the transformation of industries: three research perspectives. *The Journal of Strategic Information Systems*, *13*(1), 5–28. https://doi.org/10.1016/j.jsis.2004.02.001

Cummings, J. N., & Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks*, *25*(3), 197–210. https://doi.org/10.1016/S0378-8733(02)00049-7

Dalton, D. R., & Todor, W. D. (1979). Turnover Turned over: An Expanded and Positive Perspective. *The Academy of Management Review*, *4*(2), 225. https://doi.org/10.2307/257776

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: the New Science of Winning*. Harvard Business School Press.

de Oliveira, J. M., Zylka, M. P., Gloor, P. A., & Joshi, T. (2019). Mirror, Mirror on the Wall, Who Is Leaving of

Them All: Predictions for Employee Turnover with Gated Recurrent Neural Networks. In *Collaborative Innovation Networks* (pp. 43–59). Springer, Cham. https://doi.org/10.1007/978-3-030-17238-1_2

Deloitte. (2020). Using network analysis to intentionally architect collaboration (Issue October).

DeSanctis, G. (1986). Human Resource Information Systems: A Current Assessment. *MIS Quarterly*, *10*(1), 15. https://doi.org/10.2307/248875

Dess, G. G., & Shaw, J. D. (2001). Voluntary Turnover, Social Capital, and Organizational Performance. *Academy of Management Review*, *26*(3), 446–456.

Efrati, A., & Tam, P.-W. (2010). *Google Battles to Keep Talent*. Wall Street Journal. http://www.wsj.com/articles/SB10001424052748704804504575606871487743724

Etzioni, A. (1960). Two Approaches to Organizational Analysis: A Critique and a Suggestion. *Administrative Science Quarterly*, *5*(2), 257. https://doi.org/10.2307/2390780

Everett, M. G., & Valente, T. W. (2016). Bridging, brokerage and betweenness. *Social Networks*, *44*, 202–208. https://doi.org/10.1016/j.socnet.2015.09.001

Fernandez, V., & Gallardo-Gallardo, E. (2021). Tackling the HR digitalization challenge: key factors and barriers to HR analytics adoption. *Competitiveness Review: An International Business Journal*, *31*(1), 162–187. https://doi.org/10.1108/CR-12-2019-0163

Fischbach, K., Schoder, D., & Gloor, P. a. (2008). Analyse informeller Kommunikationsnetzwerke am Beispiel einer Fallstudie. *Wirtschaftsinformatik*, *51*(2), 164–174. https://doi.org/10.1007/s11576-008-0124-z

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35–41. http://www.jstor.org/stable/10.2307/3033543

Freeman, L. C. (2004). The Development of Social Network Analysis: A Study in the Sociology of Science. Empirical Press.

Gal, U., Jensen, T. B., & Stein, M.-K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization*, *30*(2), 100301. https://doi.org/10.1016/j.infoandorg.2020.100301

Galaskiewicz, J., & Wasserman, S. (1994). Introduction: advances in the social and behavioral sciences from social network analysis. In S. Wasserman & J. Galaskiewicz (Eds.), *Advances in Social Network Analysis* (pp. xi–xvi). Sage Publications.

Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces*, *4*(3), 28–36.

George, G., & Bock, A. J. (2011). The Business Model in Practice and its Implications for Entrepreneurship Research. *Entrepreneurship Theory and Practice*, *35*(1), 83–111. https://doi.org/10.1111/j.1540-6520.2010.00424.x

Gers, F. (2001). Long Short-Term Memory in Recurrent Neural Networks. *Lausanne, EPFL*, *2366*, 102. https://doi.org/10.1162/neco.1997.9.8.1735

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410–435. https://doi.org/10.1080/0960085X.2021.1927213

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Gloor, P. A. (2017). Sociometrics and Human Relationships: Analyzing Social Networks to Manage Brands, Predict Trends, and Improve Organizational Performance. Emerald Publishing Limited.

Gloor, P. A., Almozlino, A., Inbar, O., Lo, W., & Provost, S. (2014). *Measuring Team Creativity Through Longitudinal Social Signals*. http://arxiv.org/abs/1407.0440

Gloor, P. A., Fronzetti Colladon, A., & Grippa, F. (2020). The digital footprint of innovators: using email to detect the most creative people in your organization. *Journal of Business Research*, in press.

Gloor, P. A., Fronzetti Colladon, A., Grippa, F., & Giacomelli, G. (2017). Forecasting managerial turnover through e-mail based social network analysis. *Computers in Human Behavior*, *71*, 343–352. https://doi.org/10.1016/j.chb.2017.02.017

Gloor, P. A., Oster, D., & Fischbach, K. (2012). JazzFlow—Analyzing "Group Flow" Among Jazz Musicians Through "Honest Signals." *KI - Künstliche Intelligenz*, *27*(1), 37–43. https://doi.org/10.1007/s13218-012-0230-3

Gloor, P. A., Zylka, M. P., Fronzetti Colladon, A., & Makai, M. (2022). 'Entanglement' – A new dynamic metric to measure team flow. *Social Networks*, *70*, 100–111. https://doi.org/10.1016/j.socnet.2021.11.010

Gnyawali, D. R., & Madhavan, R. (2001). Cooperative Networks and Competitive Dynamics: A Structural Embeddedness Perspective. *Academy of Management Review*, *26*(3), 431–445.

Goerzen, A. (2007). Alliance networks and firm performance: The impact of repeated partnerships. *Strategic Management Journal*, *28*(5), 487–509. https://doi.org/10.1002/smj.588

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Hasan, S. (2012). Group based trajectories of network formation and dynamics. *Social Networks*, *34*(4), 506–514. https://doi.org/10.1016/j.socnet.2012.04.001

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1–32. https://doi.org/10.1162/neco.1997.9.8.1735

Hofer, C., Barker, J. M., D'Oria, L., & Johnson, J. L. (2022). Broadening our understanding of interfirm rivalry: A call for research on how supply networks shape competitive behavior and performance. *Journal of Supply Chain Management*, *58*(2), 8–25. https://doi.org/10.1111/jscm.12281

Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). 5 Turnover and Retention Research: A Glance at the Past, a Closer Review of the Present, and a Venture into the Future. *The Academy of Management Annals*, *2*(1), 231–274. https://doi.org/10.1080/19416520802211552

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology.*

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, *12*(12), 767–797. https://doi.org/10.17705/1jais.00282

Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the HRM special issue. *Human Resource Management*, *57*(3), 679–684. https://doi.org/10.1002/hrm.21916

Isson, J. P., & Harriott, J. S. (2016). People analytics in the era of big data: Changing the way you attract, acquire, develop, and retain talent. Wiley & Sons.

Karmarkar, U. S., & Apte, U. M. (2007). Operations management in the information economy: Information products, processes, and chains. *Journal of Operations Management*, *25*(2), 438–453. https://doi.org/10.1016/j.jom.2006.11.001

Kilduff, M., & Tsai, W. (2003). *Social networks and organizations*. Sage Publications.

King, K. G. (2016). Data Analytics in Human Resources. *Human Resource Development Review*, *15*(4), 487–495. https://doi.org/10.1177/1534484316675818

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

Kitts, J. A., Lomi, A., Mascia, D., Pallotti, F., & Quintane, E. (2017). Investigating the Temporal Dynamics of Interorganizational Exchange: Patient Transfers among Italian Hospitals. *American Journal of Sociology*, *123*(3), 850–910. https://doi.org/10.1086/693704

Kleinberg, J. M. (2000). Navigation in a small world. *Nature*, *406*(6798), 845–845. https://doi.org/10.1038/35022643

Krackhardt, D., & Porter, L. W. (1986). The snowball effect: Turnover embedded in communication networks. *Journal of Applied Psychology*, *71*(1), 50–55. https://doi.org/10.1037/0021-9010.71.1.50

Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M., & Russell, G. (2018). Analytical abilities and the performance of HR professionals. *Human Resource Management*, *57*(3), 715–738. https://doi.org/10.1002/hrm.21854

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, *80*(5), 056117. https://doi.org/10.1103/PhysRevE.80.056117

Laumer, S., Maier, C., & Weitzel, T. (2022). HR machine learning in recruiting. In S. Stohmeier (Ed.), *Handbook of Research on Artificial Intelligence in Human Resource Management* (pp. 105–126). Edward Elgar Publishing.

Lazear, E. P. (2004). Balanced Skills and Entrepreneurship. *The American Economic Review*, *94*(2).

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lee, T. W., Hom, P. W., Eberly, M. B., & Mitchell, T. R. (2017). On the Next Decade of Research in Voluntary Employee Turnover. *Academy of Management Perspectives*, *31*(3), 201–221. https://doi.org/10.5465/amp.2016.0123

Leonardi, P., & Contractor, N. S. (2018). Better People Analytics: Measure Who They Know, Not Just Who They Are. *Harvard Business Review*, *November-December*, 22. https://hbr.org/2018/11/better-people-analytics

Lichtman, M. (2013). Qualitative Research in Education: A User's Guide. Sage Publications.

Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). The Analytics Journey. *Analytics Magazine*, 11–18. https://doi.org/https://doi.org/10.1287/LYTX.2010.06.01

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, *28*(1), 3–26. https://doi.org/10.1080/09585192.2016.1244699

McGrath, R. G. (2010). Business Models: A Discovery Driven Approach. *Long Range Planning*, *43*(2–3), 247–261. https://doi.org/10.1016/j.lrp.2009.07.005

McKinsey Global Institute. (2012). *The social economy: Unlocking value and productivity through social technologies*. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy

Mead, G. H. (1934). *Mind, self and society*. University of Chicago Press.

Mehra, A., Kilduff, M., & Brass, D. J. (2001). The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance. *Administrative Science Quarterly*, *46*(1), 121–146.

Mehra, A., Smith, B. R., Dixon, A. L., & Robertson, B. (2006). Distributed leadership in teams: The network of leadership perceptions and team performance. *The Leadership Quarterly*, *17*(3), 232–245. https://doi.org/10.1016/j.leaqua.2006.02.003

Minbaeva, D. (2017). Human capital analytics: why aren't we there? Introduction to the special issue. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 110–118. https://doi.org/10.1108/JOEPP-04-2017-0035

Mobley, W. H. (1982). Some Unanswered Questions in Turnover and Withdrawal Research. *Academy of Management Review*, *7*(1), 111–116.

Moody, J., & White, D. R. (2003). Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups. *American Sociological Review*, *68*(1), 103. https://doi.org/10.2307/3088904

Moreno, J. L. (1937). Sociometry in Relation to Other Social Sciences. *Sociometry*, *1*(1/2), 206–219.

Muchinsky, P. M., & Morrow, P. C. (1980). A multidisciplinary model of voluntary employee turnover. *Journal of Vocational Behavior*, *17*(3), 263–290. https://doi.org/10.1016/0001-8791(80)90022-6

Mukherjee, S. (2016). Leadership network and team performance in interactive contests. *Social Networks*, *47*, 85–92. https://doi.org/10.1016/j.socnet.2016.05.003

Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, *89*(20), 208701. https://doi.org/10.1103/PhysRevLett.89.208701

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37.

Oinas-Kukkonen, H., Lyytinen, K., & Yoo, Y. (2010). Social Networks and Information Systems: Ongoing and Future Research Streams. *Journal of the Association for Information Systems*, *11*(2), 61–68.

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems*, *25*(6), 493–508. https://doi.org/10.1057/s41303-016-0020-3

Park, H., Han, S. H., Rojas, E. M., Son, J., & Jung, W. (2011). Social Network Analysis of Collaborative Ventures for Overseas Construction Projects. *Journal of Construction Engineering and Management*, *137*(5), 344–355. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000301

Patrashkova-Volzdoska, R. R., McComb, S. A., Green, S. G., & Compton, W. D. (2003). Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. *IEEE Transactions on Engineering Management*, *50*(3), 262–269. https://doi.org/10.1109/TEM.2003.817298

Podolny, J. M. (1993). A Status-Based Model of Market Competition. *American Journal of Sociology*, *98*(4), 829–872. https://doi.org/10.1086/230091

Podolny, J. M., Stuart, T. E., & Hannan, M. T. (1996). Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991. *American Journal of Sociology*, *102*(3), 659–689. https://doi.org/10.1086/230994

Porac, J. F., Thomas, H., Wilson, F., Paton, D., & Kanfer, A. (1995). Rivalry and the Industry Model of Scottish Knitwear Producers. *Administrative Science Quarterly*, *40*(2), 203. https://doi.org/10.2307/2393636

Presthus, R. (1962). *The organizational society: An analysis and a theory.* Vintage Books/Random House. https://doi.org/10.1037/14385-000

Price, J. L. (1977). *The study of turnover.* Iowa State University Press.

Quinn, R. W. (2005). Flow in Knowledge Work: High Performance Experience in the Design of National Security Technology. *Administrative Science Quarterly*, *50*(4), 610–641. https://doi.org/10.2189/asqu.50.4.610

Raz, O., Gloor, P. a., & Survival, S. (2007). Size Really Matters: New Insights for Start-Ups' Survival. *Management Science*, *53*(2), 169–177. https://doi.org/10.1287/mnsc.1060.0609

Reinholt, M., Pedersen, T., & Foss, N. J. (2011). Why a Central Network Position Isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. *Academy of Management Journal*, *54*(6), 1277–1297. https://doi.org/10.5465/amj.2009.0007

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Schafheitle, S. D., Weibel, A., Ebert, I. L., Kasper, G., Schank, C., & Leicht-Deobald, U. (2020). No stone left unturned? Towards a framework for the impact of datafication technologies on organizational control. *Academy of Management Discoveries*, amd.2019.0002. https://doi.org/10.5465/amd.2019.0002

Schiemann, W. A., Seibert, J. H., & Blankenship, M. H. (2018). Putting human capital analytics to work: Predicting and driving business success. *Human Resource Management*, *57*(3), 795–807. https://doi.org/10.1002/hrm.21843

Schreiber, R. R., & Zylka, M. P. (2020). Social Network Analysis in Software Development Projects: A Systematic Literature Review. *International Journal of Software Engineering and Knowledge Engineering*, *30*(03), 321–362. https://doi.org/10.1142/S021819402050014X

Scott, J. (2000). *Social Network Analysis: A Handbook*. Sage Publications.

Seddon, P. B., & Scheepers, R. (2011). Towards the improved treatment of generalization of knowledge claims in IS research: drawing general conclusions from samples. *European Journal of Information Systems*, *21*(1), 6–21. https://doi.org/10.1057/ejis.2011.9

Shadish, W. R., Cook, T. D., & Campbe, D. T. (2001). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Wadsworth Publishing.

Sharma, A., & Sharma, T. (2017). HR analytics and performance appraisal system. *Management Research Review*, *40*(6), 684–697. https://doi.org/10.1108/MRR-04-2016-0084

Simmel, G. (1908). *Soziologie*. Duncker and Humblot.

Soda, G., Mannucci, P. V., & Burt, R. S. (2021). Networks, Creativity, and Time: Staying Creative through Brokerage and Network Rejuvenation. *Academy of Management Journal*, *64*(4), 1164–1190. https://doi.org/10.5465/amj.2019.1209

Somaya, D., Williamson, I. O., & Lorinkova, N. (2008). Gone But Not Lost: The Different Performance Impacts of Employee Mobility Between Cooperators Versus Competitors. *Academy of Management Journal*, *51*(5), 936–953. https://doi.org/10.5465/AMJ.2008.34789660

Sparrowe, R. T. ., Liden, R. C. ., Wayne, S. J. ., & Kraimer, M. L. (2001). Social Networks and the Performance of Individuals and Groups. *Academy of Management Journal*, *44*(2), 316–325. https://doi.org/10.5465/3069458

Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Fischbach, K., Schlagwein, D., & Schoder, D. (2014). Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors. *Academy of Management Proceedings*, *2014*(1), 15984. https://doi.org/10.5465/ambpp.2014.15984abstract

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., & Schoder, D. (2015). Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study. *Information Systems Journal*, n/a-n/a. https://doi.org/10.1111/isj.12073

Staw, B. M. (1980). The Consequences of Turnover. *Journal of Occupational Behavior*, *1*(4), 253–273.

Steel, P. (2002). At the Empirical Turnover Theory of Fit Interface: Problems. *Academy of Management Review*, *27*(3), 346–360. https://doi.org/10.5465/amr.2002.7389900

Stone, D. L., Deadrick, D. L., Lukaszewski, K. M., & Johnson, R. (2015). The influence of technology on the future of human resource management. *Human Resource Management Review*, *25*(2), 216–231. https://doi.org/10.1016/j.hrmr.2015.01.002

Tannenbaum, S. I. (1990). HRIS: user group implications. *Journal of Systems Management*, *41*(1), 27–32, 36.

Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, *43*(2–3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management*, *43*, 224–247. https://doi.org/10.1016/j.ijinfomgt.2018.08.002

Tursunbayeva, A., Pagliari, C., Di Lauro, S., & Antonelli, G. (2022). The ethics of people analytics: risks, opportunities and recommendations. *Personnel Review*, *51*(3), 900–921. https://doi.org/10.1108/PR-12-2019-0680

Uzzi, B. (1996). The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. *American Sociological Review*, *61*(4), 674–698.

Valente, T. W. (2012). Network Interventions. *Science*, *337*(6090), 49–53. https://doi.org/10.1126/science.1217330

Veit, D., Clemons, E. K., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J. M., Loos, P., & Spann, M. (2014). Business Models – An Information Systems Research Agenda. *Business & Information Systems Engineering*, 1–15.

Wang, L., & Cotton, R. (2018). Beyond Moneyball to social capital inside and out: The value of differentiated workforce experience ties to performance. *Human Resource Management*, *57*(3), 761–780. https://doi.org/10.1002/hrm.21856

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wen, Q., Gloor, P., Fronzetti Colladon, A., Tickoo, P., & Joshi, T. (2019). Finding top performers through email patterns analysis. *Journal of Information Science*, in press. https://doi.org/10.1177/0165551519849519

Wirtky, T., Laumer, S., Eckhardt, A., & Weitzel, T. (2016). On the Untapped Value of e-HRM – a Literature Review. *Communications of the Association for Information Systems*, *38*, 20–83.

https://doi.org/10.17705/1CAIS.03802

Woehler, M., Floyd, T. M., Shah, N., Marineau, J. E., Sung, W., Grosser, T. J., Fagan, J., & Labianca, G. (Joe). (2021). Turnover during a corporate merger: How workplace network change influences staying. *Journal of Applied Psychology*, *106*(12), 1939–1949. https://doi.org/10.1037/apl0000864

Xue, L., Management, I., Ray, G., Sambamurthy, V., & Lansing, E. (2011). Efficiency or Innovation: How do Industry Environments Moderate the Effects of Firms' IT Asset Portfolios? *MIS Quarterly*, *48824*.

Zott, C., Amit, R., & Massa, L. (2011). The Business Model: Recent Developments and Future Research. *Journal of Management*, *37*(4), 1019–1042. https://doi.org/10.1177/0149206311406265

Zylka, M. P., & Fischbach, K. (2017). Turning the Spotlight on the Consequences of Individual IT Turnover. *ACM SIGMIS Database*, *48*(2), 52–78. https://doi.org/10.1145/3084179.3084185

Appendix

Paper I

Bibliographic data	Schreiber, R. R., Zylka, M. P. (2020). *Social Network Analysis in Software Development Projects: A Systematic Literature Review.* In: International Journal of Software Engineering and Knowledge Engineering, Vol. 30, No. 3, pp. 321-361, https://doi.org/10.1142/S021819402050014X
Status of publication	Published as completed research paper.
Publication type	Journal article
Version & Copyright notice	Accepted version Electronic version of an article published as International Journal of Software Engineering and Knowledge Engineering, Vol. 30, No. 3, 2020, pp. 321-361 https://doi.org/10.1142/S021819402050014X © World Scientific Publishing Company https://www.worldscientific.com/worldscinet/ijseke
Previous versions	n/a

Social Network Analysis in Software Development Projects: A Systematic Literature Review

Software development in project teams has become more and more complex, with increasing demands for information and decision making. Software development in projects also hugely depends on effective interaction between people, and human factors have been identified as key to successful software projects. Especially in this context, managing and analyzing social networks is highly important. The instrument of social network analysis (SNA) provides fine-grained methods for analyzing social networks in project teams, going beyond the traditional tools and techniques of project management. This article examines the importance of the application of SNA in software development projects. We conducted a systematic literature review (SLR) of research on software development projects and social network data published between 1980 and 2019. We identified and analyzed 86 relevant studies, finding that research on software development projects spans the topics of project organization, communication management, knowledge management, version and configuration management, requirement management, and risk management. Further, we show that most studies focus on project organization and that the most common method used to gather social data relies on automated extraction from various software development repositories in the SNA context. Our paper contributes to the software development literature by providing a broad overview of published studies on the use of social networks in helping software development projects. Finally, we identify research opportunities and make suggestions for addressing existing research gaps.

Keywords: Social networks, Social networks analysis, Software development, Project management, Systematic literature review

Introduction

Creating unique software solutions gives modern enterprises an edge in the competition to develop essential product and process innovations to achieve sustainable business benefits [16, 62, 96]. In addition, software organizations are continually challenged by the need to improve the quality of software products [41]. Therefore, the importance of software development projects has grown tremendously over the last several years [65, 102].

The software development process has become more and more complex, with increasing demands for information, knowledge, decisions, and teamwork. Software development projects face several major challenges, including long geographic distances between participants across different time zones, cooperation at the company level, and open software development standards [1, 66, 108, 151]. Furthermore, the importance of networking, scalability, and open standards continues to rise [55]. An additional challenge is the participatory development that becomes more and more important due to increasing complexity of software projects [20, 29, 147]. These software developments in projects also hugely depend on effective interaction between people, and the human factors have been identified as key to successful software projects [3, 9, 28, 44, 66, 107, 153].

Despite the standardization of software development through process models [46, 97], the success of effective and efficient software development continues to be influenced significantly by social networks. One of the most common causes for software development project failure is poor communication and interaction among customers, developers, users, and project managers [25]. These actors are embedded in social networks that need to be developed and managed purposefully to enhance the effectiveness and efficiency of software development [113]. In this context, managing and analyzing social networks is highly important. In particular, social network analysis (SNA) in software project development projects has the potential to reveal hidden structural issues and collaboration patterns [52].

Prior research on software development projects has focused on open-source software (OSS). OSS projects are self-organized and dynamic processes in which volunteers around the world contributes to a software product [57, 130]. The source code in open-source projects is made available with a license from the copyright holder that provides the rights to study, change, and distribute the software to anyone and for any purpose [86]. In contrast, the generated source code in closed-source development projects is not made freely available to the public [32]. Mixed business models provide open-source code and offer commercial support [27].

While research focused on social networks in OSS projects has increased, there are fewer research studies addressing the importance and influence of social networks in closed-source software development projects. In contrast to OSS projects, enterprises can select project members from among their own internal employees or from external resources, such as partners [67]. Furthermore, in enterprise-managed software development projects, there is generally more flexibility in defining organizational structure, spatial distribution, task allocation, and competence distribution compared to OSS projects [113]. However, while an enterprise-managed project's social network may be better planned, designed, and controlled than that of an OSS development project, research on OSS projects' social networks suggests they may have similar effects on closed-source enterprise-led software development projects [11].

The purpose of this study is to provide an overview of current SNA applications in software development projects. Further, we propose a research agenda for investigating software development projects with a focus on social networks based on a systematic literature review (SLR) using Kitchenham and Charters (2007) approach. As our paper shows, the results of this literature review reveal new and unexplored aspects of software development projects with respect to project organization, communication management, knowledge management, version and configuration management, requirement management, and risk management. This article provides a theoretical basis to formulate recommendations for future research. The results of this paper are particularly relevant for researchers seeking an overview of the topic, insights into software development projects and the role of social networks, and opportunities to contribute to research in this specialized field.

The remainder of the article is organized as follows. In the next section, we provide brief background information on SNA and software development projects, the two key concepts in our literature review. Section 3 introduces the literature review, including the research method we used to conduct the study. In Section 4, we analyze the literature and answer relevant research questions. Section 5 discusses our limitations, and we conclude in Section 6 by highlighting the paper's contributions and offering suggestions for future research.

Background

Social Network Analysis

The subject of our research is social networks within software development projects. These projects are complex systems and can be analyzed using SNA, an interdisciplinary approach combining research in anthropology, sociology, psychology, and organizational theory [14, 38, 103]. SNA places people and their relationships at the center of research activities and provides systematic methods and measures to identify, visualize, and analyze social structures and orders [103]. Many studies have demonstrated SNA's usefulness in the business context (e.g., to support strategy collaboration [37]) and in the organizational context (e.g., for innovation support and knowledge work [38, 103]).

Social networks are described through rational, structural, and functional characteristics. The rational dimension focuses on the links between pairs of individuals and can be described in terms of, for example, intensity, reciprocity, and multiplexity. Structural properties, such as density and the size of a network, correspond to the morphology of relationships between actors. Functional dimensions refer to transactional content, such as how two workers might exchange knowledge or information.

Table 1 lists the basic social network properties used in our investigation and their different dimensions that we encountered during our systematic literature review.

Structural characteristics		
Property	**Definition**	**Example Interpretation**
Density	The number of direct ties in a network as a ratio of the total number of possible links [142].	Well-connected software development teams have many direct ties between team members. A low density indicates a software developer's little development involvement in the project activities.
Size	The sum of relevant actors [142].	Thus, network size metric allowing comparisons of actors across different projects. A software project with a high number of actors has a high significance.
Bridge	An individual node that provides the only link between multiple clusters [134].	Team member that connecting separate project teams are a bridge. Bridges are important channels for information transitions in software projects. Two developers cannot further interact with each other when the bridge is broken.
Degree centrality	The number of direct ties to other nodes [121].	Influence, popularity, important software developer in teams tend to have a high degree

		centrality. If a developer is very popular, he has many ties to other actors in the social network.
Closeness centrality	A measure calculated as the sum of the length of the shortest paths between a node and all other nodes in a social network [54].	A high closeness centrality metric shows a rapid information distribution in organizations. If the closeness centrality for one software developer is high, it could be the most important influencer.
Betweenness centrality	The number of shortest paths that pass through a given node [54].	This measure shows which nodes act as 'mediator' between nodes in a social network. Manager with high betweenness centrality can control the communication flow around a software project.
Rational characteristics		
Concept	**Definition**	**Example Interpretation**
Reciprocity	A measure of mutually linked nodes [142].	Connected software developer in software projects that help each other's in different problem situations.
Multiplexity	The number of content forms contained in a tie between nodes [134].	Software project member interact in multiple contexts like coworkers and advisors. The analysis of multiplex networks provides a more realistic and more detailed picture of these social networks.
Homogeneity	The similarity of nodes [120].	The node actor homogeneity refers to similarities of actors with respect to e.g. age, country, and speech. The members of software development projects could be with knowledge of the same programming language and speech.
Intensity	The strength of the relationship between nodes in a network [121].	Software developers that communicate with each other often have high intensity. Intense ties tend to be situated within subgroups in large projects.
Functional characteristics		
Transactional content	What is exchanged through the link of two nodes [134].	The relations between software developers can have different contents and implications. Examples are friendship, information, or services that are exchanged between project members.

Table 1. Network properties

Many social network theories are based on these basic social network properties and provide additional perspectives on complex social ties. We identified eight theories for our literature review, which are summarized in Table 2. These theories are useful and relevant for understanding social networks in software development projects.

Theory concepts	Definition	Example Interpretation
Clique analysis	This analysis of social structure focuses on how connections between large social structures can be built with small and tight components [78].	Core-periphery project structure of software development teams. The importance of having a stable and healthy core team is essential to software projects.
Embeddedness	Trusting relationships between actors tend to expand through broker exchange. Trust acts as the primary governance structure in cooperation (structural effect of transitivity) [137].	The contexts and relationships in which software developers are embedded influence the information and knowledge they receive.
Structural holes	Individuals hold certain positional advantages or disadvantages based on how they are embedded in social structures. A structural hole is a gap between two nodes with complementary sources of information [18].	Identify developer who connects different groups to fill these structural holes. Structural holes emphasize a team member's abilities to access a wide range of information, resources, or perspectives as a key factor for efficient team performance.
Strength of weak ties	This theory proposes that the most important links in a network are not the strongest connections but the weaker connections that link otherwise unconnected groups through a network bridge [59].	Weak ties of software developers provide access to novel information from other projects. In particular, weak ties provide access to nonredundant information.
Social capital	Social capital is the sum of the actual and potential resources embedded within, available through, and derived from an individual's or social unit's network of relationships [104].	Developer's access to resources from a network of relationships, may emerge as a key factor that differentiates those who are more productive than others.
Power-law distribution	A few actors have many incoming social ties, and predominant actors have just a few ties [6].	Social network in which a small fraction of the software developers occupies many links while most of the nodes only have a few links.
Conway's Law	This law states that organizations design systems that typically reflect their own communication structures [34].	The software design typically reflects the organizational structure of its development team. This can generate dependencies between different software developers.
Small world	Small-world experiments examine the shortest average path length for social networks [143].	Structure of relationships among developers affects how efficient information and knowledge is shared among them.

Table 2. SNA concepts

Software development projects

The software development process comprises highly complex tasks and subtasks that cover different aspects of software development and project management. A variety of conceptual models exists that aim to structure and define these tasks and subtasks to make them more manageable and provide guidance regarding the order in which a project's tasks should be carried out [12]. Each model has its own strengths and weaknesses and is not necessarily suitable for all types of software development projects [12]. Complexity is a major influencing factor when it comes to choosing an appropriate software development model. Hence, some past software development models, such as the waterfall model [115], spiral model [12], rapid prototyping model [56], and several others, are either sequential or cyclical. They all cover the same important phases of software development: (1) planning and analysis, (2) design, (3) implementation, and (4) testing. These phases are also covered by more general models, such as the systems development life cycle (SDLC) [111, 118] and Standard ISO/IEC 12207 "Information Technology Software Lifecycle Processes" [73]. Every software development project management approach uses this generalized four-phase method.

The four phases cover six major classes of management topics related to software development projects, as shown in Table 3. Each of these topics corresponds to a management process and fulfills an overarching role in the course of software development projects, making it crucial for project management. Figure 1 shows how the corresponding processes fit into this context.

Topic	**Description**
Project organization (PO)	Project organization includes the organizational and operational structure of a project. The organizational structure is set within the company to identify and define individual work processes [17].
Communication management (CM)	Essential tasks of communication management include planning, collecting, distributing, and storing project information [88].
Knowledge management (KM)	Knowledge management involves activities and processes for leveraging knowledge to enhance competitiveness through the creation and use of individual and collective knowledge resources [89, 122].
Version and configuration management (VM)	Software projects create many artifacts (i.e., project results). These documents, databases, programs, and pieces of source code typically exist in different versions and must be managed accordingly. Software configuration management tasks include identifying, organizing, and supervising changes to artifacts [64].
Requirement management (RM)	The requirement management process describes activities for obtaining and controlling requirement changes and ensures that related to relevant plans and data are kept up to date. Further, it provides requirements traceability and ensures that changes to requirements are reflected in project plans, activities, and work products [30].
Risk management (RI)	The focus of risk management is to identify potential uncertainties and evaluate and derive appropriate actions. The aim is to minimize project risks [51, 83].

Table 3: Management topics in software development projects

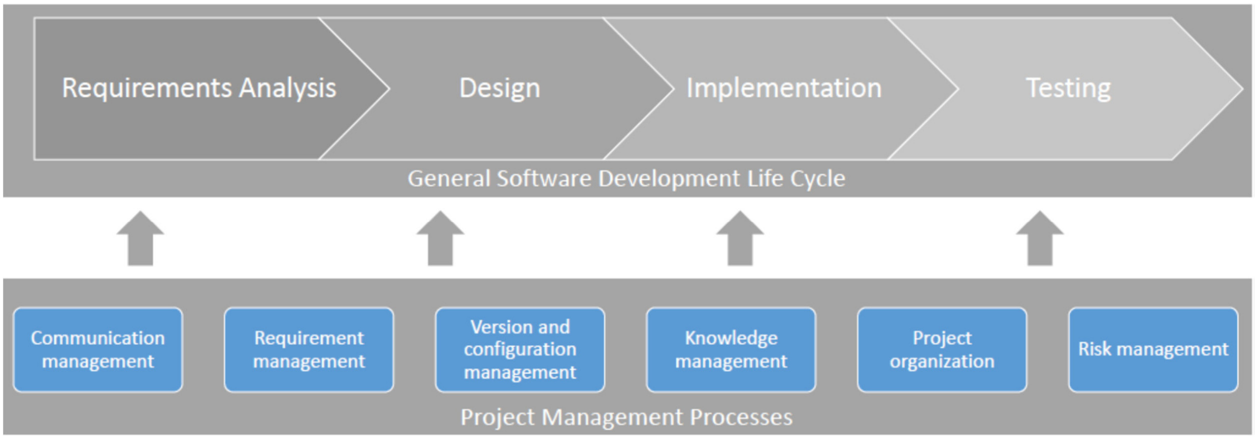


Figure 1: Overview of software development project management processes

Research method

To investigate potential research gaps and the potential for using social network analysis in software development projects, we conducted an SLR, a well-structured and repeatable method to identify, evaluate, and interpret existing literature relevant to a specific research topic [35, 82, 139, 144]. We followed guidelines for performing SLRs in software engineering [15, 82]. Our SLR focuses on research outcomes and methods covering two areas: software development projects and SNA. Our focus on propositions and concepts in these two research areas inform the organization of this paper, which follows a conceptual structure. Also, to guarantee a neutral representation of the review results, we adopt no particular perspective. The audience for this review is specialized scholars interested in research of using SNA in the context of software development projects. We show ways to contribute to the development of research in this specialized field.

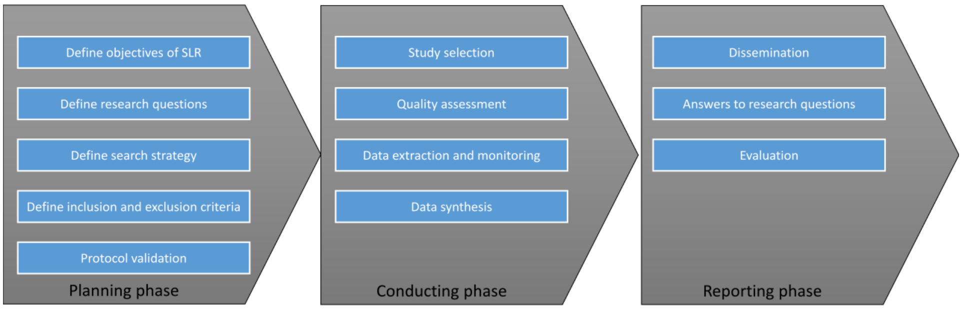


Figure 2: Main SLR phases [82]

We split our SLR into three phases, as suggested by Kitchenham and Charters (2007): planning, conducting, and reporting. These phases and their sub-steps ensure rigor and transparency throughout the SLR process (see Figure 2).

Research questions

The research questions outline what should be addressed and evaluated from the selected publications, so defining suitable research questions is the most important activity during the planning phase [82]. Table 4 presents the research questions we chose to achieve the objectives of this SLR.

No.	Research question
RQ1	Which main research trends can be identified by analyzing relevant publications about SNA in software development projects?
RQ2	What are the main research findings about SNA in software development projects?
RQ3	What future research opportunities does current research reveal?

Table 4: Research questions

Search string

To identify relevant publications, we conducted a keyword-based search using electronic databases. The keywords initially identified were derived from the research questions (see Table 5), from trial searches, from consultations with experts in the field and using the PICO method (population, intervention, comparison, outcome) [82]:

- Population:Social network analysis
- Intervention: Software project management
- *Comparison:not relevant (this is an exploratory study)*
- Outcome:Source code

To find higher quality and more relevant references from the electronic database search within the management topics in software development projects, the search terms were not chosen too restrictively [82].

Main topics	Related terms
Social network analysis	SNA Social network
Software project management	Software project Software development Project development Project governance
Source code	Closed source Open source

Table 5. Search Keywords

The resulting search terms with similar meanings were organized into logical groups. Combined terms are obtained using the OR logical operator in the same group. The final search string was obtained using the AND logical operator between the different logical groups. The complete database search string is shown in Table 6.

(“social network analysis” || “social network*”)
&
(“software project management” || “software project*” || “software development” || “project governance”)
&
(“source code” || “closed source” || “open source”)

Table 6: Search string

Inclusion and exclusion criteria

We considered all papers published from January 1980 to April 2019, excluding studies not peer reviewed as well as those that appeared as research-in-progress studies in conference proceedings.

To identify relevant publications, we conducted a keyword-based search using the following databases Table 7. The digital search libraries were selected because they were known to have been used as source for other systematic literature reviews related to software engineering [15, 112]. These sources include certain highly important journals, such as: Information and Software Technology, IEEE Software, Computer, Information and Management, and Systems and Software. Moreover, we included Google scholar, a web search engine that indexes the full text or metadata of scholarly literature across different disciplines to get an extended perspective. The efficient use of Google scholar to perform bibliometric studies has been demonstrated [114]. Nevertheless, the Google scholar restrictions must be respected in our keyword-based search [13].

Digital Libraries	URL
ScienceDirect (subject Computer Science)	https://www.sciencedirect.com/browse/journals-and-books?subject=computer-science
Wiley InterScience (subject Computer Science)	https://www.wiley.com/en-us/General+%26+Introductory+Computer+Science-c-CS00
IEEE Digital Library	https://www.computer.org/csdl
ACM Digital Library	https://dl.acm.org/
Google Scholar	https://scholar.google.de/

Table 7. Search databases

The search string presented in Table 6 was adapted to each search engine, related to the features or restrictions that each search engine had. Several databases use a different syntax e.g. for Boolean operators. In these cases, we replaced “||” with “OR” or vice versa. The command search queries used in each digital library and search engine are shown in Table 8.

Source	Specific search string	Extra options	Results
ScienceDirect	Social network software project	Using the abstract, title, and keyword search; 1980 – 2019	106
Wiley InterScience	"(“social network analysis” OR “social network*”) AND (“software project management” OR “software project*” OR “software development” OR “project governance”) AND (“source code” OR “closed source” OR “open source”)	Subject Computer Science; Journals; 1980 – 2019	495
IEEE Digital Library	social network AND software project	Full text; 1980-01 – 2019-03	549
ACM Digital Library	+"social network" +"software project"	Published since 1980	67
Google Scholar	[(“social network analysis” || “social network*”) & (“software project management” || “software project*” || “software development” || “project governance”) & (“source code” || “closed source” || “open source”)]	Respect to limitations of Google Scholar; Choose period year by year	15600

Table 8. Search queries

All search results were stored in Citavi and classified as relevant or irrelevant by the reviewer team. To determine relevance, we manually screened each stored paper’s title, abstract, and keywords. We then further excluded papers that did not match our research focus and questions.

Table 9 shows our inclusion and exclusion criteria. The inclusion criteria were applied using “AND” as logical operator and the exclusion criteria using “OR” as logical operator between them.

Inclusion criteria	Exclusion criteria
Studies published after 1980	Studies without reference to software development projects
Studies that refer primarily to SNA and in software development projects	Studies that refer to the use of social media applications (e.g., Facebook or Twitter)
Books, monographs, journal articles, and conference proceedings	Studies not peer reviewed and not in a scientific context
	Studies written in languages other than German or English
	Studies appearing as research in progress, posters, or short papers of fewer than six pages

Table 9: Inclusion and exclusion criteria

Study quality selection

Each paper was checked against the quality assessment criteria inspired by Database of Abstracts of Reviews of Effects (DARE) [24] and shown in Table 10. The five quality assessments (QAs) were performed by the first author and were verified by the second author. We verified whether the publications mentioned or discussed issues related to each of the quality metrics. Any differences of opinion between the first two authors were resolved through discussions until a consensus was established.

Is the explanation of the research aim, hypothesis, and research design understandable?
For empirical research papers:
Were the data collected in a way that addressed the research issue?
Is the research method stated explicitly?
Are limitations of this study mentioned?
Is there a clear statement of findings?
Was the study published in a recognized source?

Table 10: Quality assessments

The questions were scored as follows:

- QA1:Y (yes) +1, the research aim, hypothesis and research design are clearly defined; P (partly) +0.5, the research aim, hypothesis and research design are implicitly defined; N (no) 0, the research aim, hypothesis and research design are not defined.
- QA2: For empirical research papers:
- QA2 a): Y +1, the data collected addressed the research issue; P +0.5, the data collected address implicit the research issue; N 0, the data collected not address the research issue.
- QA2 b): Y +1, the research method is clearly stated; P +0.5, the research method is implicitly stated; N 0, the research method is not stated.
- QA3: Y +1, the limitation of this study are mentioned; P +0.5, the limitation of this study not clearly addressed; N 0, there are no limitations of this study.
- QA4: Y +1, the study has clear findings; P +0.5, the study has a statement and findings; N 0, the study has not a clear statement and findings.
- QA5: Y +1, the study was published in an accredited and stable publication source. This rating is based on the 2018 journal citations report (JCR) for journals [72] and the computer science conference ranking in Computing Research and Education (CORE) 2018 [36]; N 0, the study was published in an unranked publication source.

We excluded all papers that did not match our quality assessment with a score greater than 0 in all questions.

In addition, we screened the references of the search results for other relevant research publications by applying a manual backward search. Moreover, we applied a forward search by using the “citied by” function of Google Scholar, to identify publications that cited by the publications (see Figure 3). These search results complied with the inclusion/exclusion criteria. We marked the studies we found via the backward search with “BREF” the forward search with “FREF”.

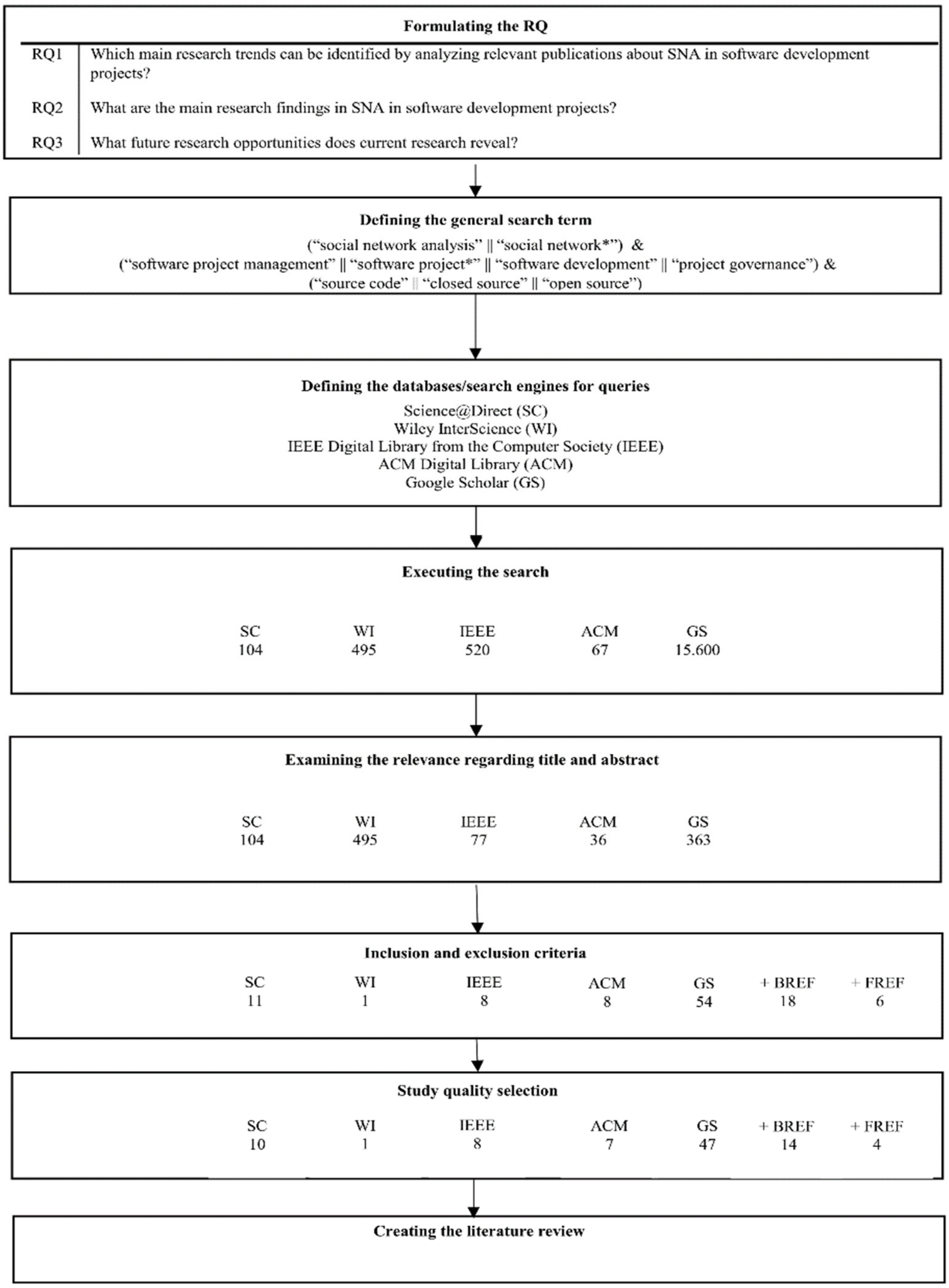


Figure 3: SLR process

The results of the search including the relevant publications are shown in the literature collection (see Appendix A). Finally, all 86 articles were read carefully for consideration for our literature review.

Data extraction strategy

Once the primary studies were chosen, the relevant data for the systematic literature review was obtained to answer our RQs. Another spreadsheet was used to extract data concerning SNA and software project management processes. The data extraction strategy was more challenging for the heterogeneous studies and is explained for each RQ here:

RQ1: Which main research trends can be identified by analyzing relevant publications in SNA software development projects?

The year, publication source, type of software source code, level of analysis, research focus in software project management process and data collection method for each paper is listed and the aggregated results will be presented.

RQ2: What are the main research findings in SNA in software development projects?

The general key findings in SNA and the results from the different research areas in software project management processes for each paper are recorded. The classification of the different processes will be based on Table 3.

RQ3: What future research opportunities does current research reveal?

The overall key research opportunities in SNA and the results from considering the different research areas will be documented.

Results

This section presents our findings from the SLR. First, we describe an overview of the relevant studies via a research matrix. Afterwards, we discuss each of the research questions in light of these findings.

Overview

In this section, we conduct an analysis of the management topics in software development projects. Additionally, we checked the relevant documents from the literature review to identify research that already exists relating to SNA in software development projects. Subsequently, the documents were classified according to the SNA level of analysis (individual, project, or organization) and the type of source code availability (closed source, open source, or both). In addition, the SNA specifics mentioned in Section 2 have been expanded in this following list. An overview of this information is presented in Table 11.

		Project organization	Communication management	Knowledge management	Version and configuration	Requirement management	Risk management	Total
Analysis level	Individual	3	0	1	0	0	0	4
	Project	43	12	13	2	2	5	77
	Organization	4	1	2	0	0	0	7
Source code availability	Both (closed and open source)	3	0	0	0	0	2	5
	Closed source	3	3	4	0	2	1	13
	Open source	43	10	11	2	0	2	68
Structural characteristics	Density	26	6	9	0	6	4	51
	Size	34	9	11	1	1	4	60
	Bridge	6	2	5	0	0	1	14
	Degree centrality	17	8	4	1	2	3	35
	Closeness centrality	5	1	2	1	2	3	14
	Betweenness centrality	14	3	3	1	2	2	25
Rational characteristics	Reciprocity	5	1	3	0	0	0	9
	Multiplexity	4	1	0	0	0	0	5
	Homogeneity	9	0	0	0	0	0	9
	Intensity	4	1	0	0	0	0	5
Functional characteristics	Transactional content	2	0	2	1	0	0	5
SNA theory concepts	Clique analysis	31	3	5	1	0	1	41
	Embeddedness	5	0	2	1	0	0	8
	Structural holes	4	1	1	0	0	0	6
	Strength of weak ties	4	0	0	0	0	0	4
	Social capital	6	0	3	0	0	0	9
	Power law distribution	6	0	1	0	0	0	7
	Conway's Law	3	1	0	0	0	0	4
	Small world	3	1	0	0	0	0	4
	Total	49	13	15	2	2	5	

Table 11: Research matrix

Main research trends

In this section, we answer our first research question, RQ1: Which main research trends can be identified by analyzing relevant publications in SNA software development projects?

Many studies focused on open-source software projects in software development projects (79.1%, 68 publications). Other publications concentrated on closed-source software projects (15.1%, 13 publications) or on both types of software projects (5.8%, five publications).

Software development projects area	Papers	Amount	Percent
Open source	S1, S2, S3, S4, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S17, S19, S22, S24, S25, S26, S27, S28, S29, S30, S31, S35, S36, S37, S39, S40, S41, S42, S43, S44, S45, S46, S47, S48, S50, S51, S52, S53, S54, S55, S57, S58, S59, S60, S61, S62, S63, S64, S65, S69, S70, S71, S72, S73, S74, S75, S76, S78, S80, S81, S83, S84, S85, S86	68	79.1%
Closed-source	S5, S20, S21, S32, S34, S38, S49, S56, S66, S67, S77, S79, S82	13	15.1%
Mixed	S16, S18, S23, S33, S68	5	5.8%

Table 12: Relative amount of publications by type of software development project

The level of analysis used in each study is also of interest, because different social network levels describe the research focus of existing researcher's investigations. The documents were classified by the SNA levels of individual, project, and organization. Project was by far the most common level of analysis (87.5%, 77 publications) followed by organization (8%, 7 publications) and individual (4.5%, four publications).

Level of analysis	Papers	Amount	Percent
Individual	S9, S17, S60, S72	4	4.5%
Project	S1, S2, S3, S4, S6, S7, S8, S10, S11, S12, S13, S15, S16, S17, S19, S20, S21, S22, S23, S24, S25, S26, S27, S29, S30, S31, S32, S33, S34, S35, S36, S37, S38, S39, S40, S42, S43, S44, S45, S47, S48, S49, S50, S51, S52, S53, S54, S55, S56, S57, S58, S59, S60, S61, S62, S64, S65, S66, S67, S68, S69, S70, S71, S73, S74, S75, S76, S77, S78, S79, S80, S81, S82, S83, S84, S85, S86	77	87.5%
Organization	S5, S14, S18, S28, S41, S46, S63	7	8.0%

Table 13: Relative amount of publications by level of analysis

There were no relevant findings of research works published before 2002 (see Table 14). Table 14 also shows the number of publications by topic, including communication management, requirement management, version and configuration management, knowledge management, project organization, and risk management, ordered by their year of publication. Project organization was the main research focus. The number of papers on the other topics was very irregular, thus preventing any major conclusions.

In general, based on the number of publications, there seems to be a volatile interest in SNA focusing on projects.

Year	Project organization	Communication Management	Knowledge Management	Version and configuration Management	Requirement Management	Risk Management	Total
2002	**1**	0	0	0	0	0	**1**
2003	0	0	0	0	0	0	**0**
2004	0	0	0	0	0	0	**0**
2005	**3**	**1**	0	0	0	0	**4**
2006	**6**	**2**	**2**	0	0	0	**10**
2007	**2**	**2**	**1**	0	0	0	**5**
2008	**3**	**2**	**1**	0	0	**1**	**7**
2009	**4**	**1**	**1**	0	0	**1**	**7**
2010	**1**	**1**	**1**	0	**1**	**1**	**5**
2011	**6**	0	**1**	0	**1**	**2**	**10**
2012	**1**	**1**	0	0	0	0	**2**
2013	**4**	0	**3**	**1**	0	0	**8**
2014	**3**	**1**	**1**	0	0	0	**5**
2015	**5**	**1**	0	0	0	0	**6**
2016	**1**	0	**2**	**1**	0	0	**4**
2017	**9**	0	**1**	0	0	0	**10**
2018	0	**1**	**1**	0	0	0	**2**

Table 14: Number of publications by topic area

Table 15 shows the research focus of the selected papers. Data on these selected publications indicated that the major research topic was project organization (57.0%, 49 publications). The other common research topics observed in our review were in the areas of knowledge management (17.4%, 15 publications) and communication management (15.1%, 13 publications).

Research areas	Papers	Amount	Percent
Project organization	S2, S5, S6, S7, S9, S10, S11, S12, S13, S17, S18, S19, S24, S25, S26, S27, S28, S30, S31, S33, S36, S37, S40, S41, S42, S43, S44, S45, S48, S50, S53, S55, S56, S57, S59, S61, S64, S65, S68, S70, S71, S72, S74, S77, S78, S81, S83, S85, S86	49	57.0%
Communication management	S1, S4, S14, S15, S20, S21, S29, S34, S47, S51, S69, S75, S84	13	15.1%
Knowledge management	S3, S8, S22, S35, S38, S39, S46, S60, S62, S63, S66, S67, S73, S76, S82	15	17.4%
Requirement management	S32, S49	2	2.3%
Risk management	S16, S23, S54, S58, S79	5	5.8%
Version and configuration management	S52, S80	2	2.3%

Table 15: Research areas

To substantiate this analysis, we looked at the social data-collection methods the relevant papers used (see Table 16). In these publications, the most common method for data collection is automated extraction from various software development repositories (80.9%, 72 publications). These databases keep track of the software development process. The second most common data-collection method observed in our review was surveys (8.0%, 8 publications).

SNA data-collection method	Papers	Amount	Percent
Content analysis from repository	S1, S2, S3, S4, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S17, S19, S21, S22, S23, S24, S25, S26, S27, S28, S29, S31, S33, S34, S35, S36, S37, S39, S40, S41, S42, S43, S44, S45, S46, S47, S48, S50, S51, S52, S53, S54, S55, S57, S58, S59, S60, S61, S63, S64, S65, S66, S69, S70, S71, S72, S73, S74, S75, S76, S77, S78, S79, S80, S81, S83, S85, S86	72	80.9%
Surveys	S18, S20, S32, S42, S49, S56, S57, S72	8	8.0%
Combinations	S5, S30, S38, S62, S67, S68, S82	7	7.9%
Simulations	S16	1	1.1%
Observations	S48	1	1.1%
None	-	0	0.0%

Table 16: SNA data-collection methods

Another major factor regarding influential publications is the citation counts from the digital library Google Scholar. We took the top 10 research papers with the most citations. The search was executed in April 2019 and is shown in Table 17.

Title	Paper	Research area	Year	Citations
Mining email social networks	S51	Communication Management	2006	596
The social structure of Free and Open Source software development	S69	Communication Management	2015	543
Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems	S70	Project organization	2006	505
Socialization in an Open Source Software community: A socio-technical analysis.	S72	Project organization	2005	425
The Open Source Software Development Phenomenon: An Analysis based on Social Network Theory	S7	Project organization	2002	327
Latent Social Structure in Open Source Projects	S50	Project organization	2008	269
Predicting Failures with Developer Networks and Social Network Analysis	S79	Risk Management	2008	217
Social dynamics of free and open source team communications	S29	Communication Management	2006	186

Hierarchy and Centralization in Free and Open Source Software Team Communications	S71	Project organization	2006	186
Network effects: The influence of structural capital on open source project success	S13	Project organization	2011	156

Table 17: Influential research publications

Table 18 shows the top 10 important sources and channels of publications for SNA in software development project research. This result shows that 8 selected papers were presented at the Hawaii International Conference on System Sciences (HICSS) and its related workshops. This analysis shows the main publication sources and the journal or conference rank as a quality indicator.

Publication source	**Publication channel**	**Ranking**	**Papers**	**Amount**
Hawaii International Conference on System Sciences (HICSS)	Conference	CORE A	S15, S17, S19, S35, S46, S53, S56, S59	8
International Conference on Software Engineering (ICSE)	Conference	CORE A*	S30, S32, S42, S57, S82, S80	6
International Federation for Information Processing (IFIP)	Conference	CORE A	S3, S25, S26, S29, S40, S83	6
Information and Software Technology (IST)	Journal	JCR 2.694	S38, S52, S60, S62, S66	5
International Conference on Information Systems (ICIS)	Conference	CORE C	S2, S4, S28, S44	4
Empirical Software Engineering	Journal	JCR 3.275	S43, S74	2
First Monday	Journal	JCR 0.563	S12, S69	2
Journal of Software and Systems (JSS)	Journal	JCR 2.444	S47, S68	2
The Americas Conference on Information Systems (AMCIS)	Conference	CORE A	S6, S7	2
Special Interest Group on Software Engineering (SIGSOFT)	Conference	CORE B	S50, S79	2
Other	Journal	-	S8, S9, S13, S20, S36, S43, S45, S47, S61, S63, S64, S65, S67, S68, S70, S71, S72, S73, S74, S75, S78, S81, S85	23
Other	Conference	-	S1, S5, S10, S14, S16, S23, S31, S34, S37, S39, S41, S48, S49, S51, S54, S55, S58, S84	18
Other	Workshop	-	S22, S24, S77, S86	4
Other	Book	-	S11, S76	2

Table 18: Publication source and channel

Main research findings

In this section, we answer our second research question RQ2: What are the main research findings in SNA in software development projects?

There is a lot of research at the project analysis level (77 papers). Regarding source code availability, open source (68 publications) is the main focus of research on software development projects. Furthermore, the main topics in current research are project organization (49 publications) and knowledge management (15 publications).

There are different research focuses related to the SNA areas (see Table 2 - SNA theory concepts). The main SNA research topics include structural characteristics (78 studies) followed by rational characteristics (21 documents). However, many studies also concentrate on clique analysis (41 publications) and social capital (9 papers). Below, we discuss the key findings from the different research areas in software development.

Project organization. In software development projects, project organization among distributed software developer is a challenge especially to mainly geographic distances, cultural distance, different time zone and cooperation at company level [66, 151]. From the analysis, we observed that research studies in this area examined coordination and cooperation in software development projects. Research focused on social project structure and the theory concept clique analysis. This theory includes especially the topics of core periphery (core team and enhanced team) and cluster [21, 33, 75, 136]. Other results of studies in this topic suggest that there is a symmetry between social networks and design structures in different software projects. Findings support this hypothesis and confirm Conway's Law [2, 4, 84]. There are additional investigations of the influence of social capital [61, 87, 125, 126, 133, 141] and power law distribution [33, 67, 76, 95, 145, 152] in software development projects. According to previous studies, the influence of social capital on open source software projects' success and developer productivity is examined [141]. We found evidence that cohesive social ties between team members in their social networks lead to more productivity [87, 126]. It was also evidenced that the power law distribution in open source projects exists [33, 67, 76, 95, 145]. In this area, many more SNA studies have focused on the rational characteristics of social ties.

Communication management. Communication within projects, which has become increasingly complex, is one key to successful projects [106, 148, 150]. We identified many studies that use SNA to visualize and examine the communication structure between team members [39, 106]. Research in communication networks in large commercial software project found that individual software developers perform better when they are central in a team's communication network [22]. Furthermore, individuals embedded in dense communication clusters at the team and project levels perform better than those who are not embedded [48]. Another paper examined how communication develops between individuals and groups over time and their correlation to project development [58]. The method of SNA to visualize communication patterns helps teams and managers to track the progress, improve the communication and avoid communication gaps in software development projects [128]. In regard to SNA theory, there are many further investigations of clique structure in the area communication management [22, 58, 106]. Additionally, many researches focused on the structural characteristics of social ties.

Knowledge management. Several SNA findings demonstrated that network ties are important for knowledge flows in open-source [94, 135]. One research illustrated how to identify the man-in-the-middle (bridges) between other community members and proved the importance of this role [135]. Another studies addressed the direction of knowledge flows among different projects in social networks and its impact on project success [50, 99, 110]. Measurements from one work support the generalized reciprocity exchange theory in the area knowledge management [99]. An additional study demonstrated how network analysis techniques can be used to identify key authors and subject matter experts [45]. It was also evidenced that the identification of knowledge experts in open source software projects is possible with SNA methods [129]. Besides, that team members that are boundary spanners have extensive domain knowledge and hold key positions in the control project structure [42]. In this research area, many studies focused on the structural characteristics of social ties and on clique analysis [81, 90].

Version and configuration management. High quality bug tracking and code reviewing are regarded as the most important approach to preserving quality in software development projects [156, 157]. In research by Zanetti, Scholtes, Tessone, and Schweitzer (2013), the authors propose an efficient and practical method to identify valid bug reports that refer to an actual software bug and contain enough information to be processed. This developed method could easily be integrated into bug-tracking platforms, and its performance could analyze open-source communities using SNA [157]. Further research was conducted on automated recommendations for code reviewers to improve the development performance [156].

Requirement management. Two articles have examined stakeholder involvement and their importance for software development projects [91, 92]. The results of another study reveal that stakeholder communication and involvement have been recognized as a major success factor in commercial software development projects [91]. Additionally stakeholder communication and involvement have been recognized as a major success factor in commercial software development projects [91].

Risk management. In one study, Sureka, Goyal, and Rastogi (2011) presented a systematic approach for mining defect-tracking systems for risk, threat, and vulnerability analysis in software projects. They derive a collaboration network from a defect-tracking system and apply SNA techniques to demonstrate how important information pertaining to risk and vulnerability can be uncovered using SNA [130]. In another study, the authors proposed a new set of social network metrics for issue repositories to predict defects. The results of the experiments reveal that compared to other sets of metrics, using social network metrics for issue repositories considerably decreases high false-alarm rates [9]. In risk management, three papers focus on structural characteristics [9, 53, 130]. One study focuses on the SNA theory concept of clique analysis [130].

Future research opportunities

In this section, we address our third research question: RQ3: What future research opportunities does current research reveal?

Further research is needed at the individual level of analysis (three publications). Regarding source code availability, additional work needs to be conducted on closed-source and mixed software projects (three publications). Also, there is a significant lack of research on the functional characteristics (transactional content) of SNA as only five studies currently exist on the topic. In terms of SNA theory concepts, there is a small number of papers about Conway's Law (four publications), the strength of weak ties (four publications), and the small world examination (four publications).

The following is an outline of the research gaps identified for further investigation organized by topic.

Project organization. Further research is needed on closed-sourced and mixed software development projects to determine whether a similar (or even the same) power law relationship forms in closed-source/mixed projects as found in open-source projects. Furthermore, there is missing research on the dynamics of social networks in software development projects over time and the whole project lifecycle. In addition, there is a lack of work on the formation of subgroups and their behavior within software development communities. More and more companies participate in open source software to gain competitive advantages. This leads to interesting research fields in the collaboration of these competing companies and their software developer in these projects. Moreover, we lack insights into social dynamic network processes when individual developers with high centrality leave their projects. There are also shortcomings in knowledge with respect to social network behavior in projects when people leave/join project teams. Regarding SNA network properties, there are only two studies on functional characteristics.

Communication management. There is a considerable need for research on how to enable and use real-time monitoring tools that can analyze communication data in projects. Such work would enable project managers to optimize and fine tune a project during the project's runtime. There is also insufficient research on the dynamics of communication patterns in projects and their impact on success. To evade communication deficits, SNA could even help to avoid struggles in the planning phase. Generally, there is a lack in effects of the development of complex communication networks in time-series basis to social networks of software projects. In regard to SNA theory concepts, there is a significant research gap, as there are only a few papers on the strength of clique analysis (three publications), Conway's Law (one publication), small world (one publication), and structural holes (one publication).

Knowledge management. Additional research is needed regarding the identification, prioritization, and coordination of knowledge brokers and their importance in closed-source and mixed software projects. Automated text analysis methods—for example, on emails or messenger messages—could be helpful to analyze knowledge flows. We also have insufficient understanding about how knowledge is transferred from inside a project to a company's external social network. Moreover, it is important to optimally design the transfer of knowledge and to plan or transfer runtime-essential knowledge during the project period. This topic is gaining importance as more and more software projects are partially executed in cooperation with external companies [66, 151]. Currently, there is insufficient research on project knowledge flow and distribution throughout the entire duration of projects; further research into these dynamic knowledge processes is essential. In regard to SNA theory concepts, there are significant research gaps related to Conway's Law, the small-world phenomenon, and the strength of weak ties.

Version and configuration management. Initial research in version and configuration management can be used to automate bug categorization and distribution. Further investigations aimed at automated task division and assignment to the appropriate software developer is required; such work should use a combination of text analysis methods and SNA. Also, the research gap in SNA structural characteristics density and bridge is large because no study exists in this area. Regarding SNA theory concepts, there is currently one paper on the strength of clique analysis and one paper on embeddedness. This means there are many future research opportunities in this area.

Requirement management. SNA may help identify important brokers, stakeholders, or organizations/companies involved in a project. Research on how the flow of information must be optimally designed to make the best use of these stakeholders is missing. There is a lack of qualitative studies on stakeholder management as well as on stakeholder-specific characteristics. Furthermore, there is missing research in prioritization requirements over project time to accommodate changing scope, stakeholders, and roles. In this topic area, research on SNA theory concepts, particularly the rational and functional characteristics of social ties, is missing as there are no papers available. In addition, we did not find any research results regarding SNA theory concepts to this topic.

Risk management. A fundamental understanding of the specific use of standardized SNA metrics for risk management within software development projects is still lacking. In addition, the risk management process involves the identification, prioritization, subsequent prevention, and reduction of potential project risks, especially in the critical human social network sector of software development projects in complex environments. In terms of network properties and SNA theory concepts concerning risk management aspects, there is a large research gap as well. Furthermore, there are no studies about rational and functional characteristics, and only one study exists on the theory concept of clique analysis. Thus, there are many opportunities for future research in this research area.

Limitations

The purpose of this work is to provide an overview of social networks in software development projects. This study is based on the systematic research method explained in Section 3. Every SLR faces numerous limitations. Restrictions obviously result from the selection of papers, choice of methods and search strings, interpretation of heterogeneity, and application of results. The validity of the study is concerned with the trustworthiness of its results. The scheme of validity and threats distinguishes between construct validity, internal validity, conclusion validity and external validity [116, 146].

The construct validity refers to the extent to which the operational measures that are studied represent what is investigated according to the RQs aim. One problem in this review was the limitation of keywords searches in the selected online databases. The search keywords were generated by the PICO method and in several iterations to ensure high quality. Further, our own bias might have influenced the selection of papers. To decrease this validity, the search was executed and validated by two authors in several iterations. Another threat was the selection of digital search engines. We had reduced this risk by using five different digital libraries.

With regard to the internal validity of the study, our study selection process may have been inaccurate due to potential misinterpretation of the inclusion/exclusion criteria. Moreover, the classification and the decision to assign a study to a specific research area can be considered subjective. To minimize this limitation, the process was carried out by two authors and the others reviewed the final results. To prevent bias in the results, we tried to make the whole SLR procedure as traceable and clear as possible in this document.

The conclusion validity is a factor to draw the correct conclusion about relationships in our data and was concerned with the ability to replicate these findings. Regarding the study search and selection, we combined our database search strategy with the forward and backward snowball method. The aim of this combination was to balance the precision values of our search and minimize the factor bias. Additionally, each step of the SLR was validated and the periodic reviews carried out by the involved researchers.

External validity concerns how far the results of a study can be generalized. Many studies are conducted on small software development projects and hence generalizing them to large software development projects is not possible. In the few cases where SNA is conducted on social networks in small software development projects as well as case studies on large projects, the external validity is reasonable.

Conclusion

This study provides an SLR summarizing the current state of academic research on social networks in software development projects as well as a research agenda with focus on social networks in these development projects. We selected 86 relevant studies from 1980 to 2019 and, to analyze these documents, defined and specified three research questions. Based on our goal and this set of research questions, we extracted data and synthesized these studies. We summarized and organized the selected papers using a classification system we derived from software development models and their associated project management processes to highlight existing research focuses.

The six major SNA management topics related to software projects include project organization, communication management, knowledge management, version and configuration management, requirement management, and risk management.

Our key findings related to primary research trends in this area are as follows. Many studies have focused on open-source software development projects (79.1%), the most common target of analysis has been the project (87.5%), and the major research focus has been project organization (57%). Furthermore, there seems to be a volatile interest in SNA focusing on software projects as shown by the number of publications over the years (particularly between 2002 and 2018). In Section 4.3, we summarized the most important research findings, and identified and ordered them based on the categories we developed. In Section 4.4, we highlighted existing research gaps and areas for further investigations. Our review revealed many gaps in this research field with high importance for open source and commercial organizations.

Based on our answers to the research questions, the SNA provides a crucial set of methods for analyzing the structure, position, autonomy, and relation of individual actors in software development projects. In addition, the topics of project organization, communication management, knowledge management, version and configuration management, requirement management, and risk management provide a wide range of research opportunities for further investigation. Deeper analysis and description should be conducted at both the social network individual and organization levels of the project topics.

The utility of this study for academic lies in the assessment of gaps in this research area. This research field is relatively new, and we expect more work in the near future to enhance knowledge about SNA in software development projects and lead to greater success.

Appendix A

Literature collection

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S1	Analysis of Survival of Open Source Projects: a Social Network Perspective [150]	Communication management	2007	1	1	1	1	1	1	6
S2	Discovering Determinants of Project Participation in an Open Source Social Network [70]	Project organization	2009	1	1	1	0.5	1	1	5.5
S3	Impacts of Social Network Structure on Knowledge Sharing in Open Source Software Development Teams [94]	Knowledge management	2008	1	1	1	0.5	1	1	5.5
S4	Investigating Success of Open Source Software Projects: A Social Network Perspective [149]	Communication management	2007	1	1	1	1	1	1	6
S5	Mapping Social Network to Software Architecture to Detect Structure Clashes in Agile Software Development [4]	Project organization	2007	1	0.5	0.5	0.5	1	1	4.5
S6	Social Network Dynamics for Open Source Software Projects [93]	Project organization	2006	1	1	1	0.5	1	1	5.5
S7	The Open Source Software Development Phenomenon: An Analysis based on Social Network Theory [95]	Project organization	2002	1	1	1	1	1	1	6
S8	Direct and indirect knowledge spillovers: the "social network" of open-source projects [50]	Knowledge management	2011	1	1	1	1	0.5	1	5.5
S9	The Study of Open Source Software Collaborative User Model based on Social Network and Tag Similarity [26]	Project organization	2014	1	1	1	0.5	0.5	1	5
1S0	A Social Network Approach to Free/Open Source Software Simulation [140]	Project organization	2005	1	1	1	0.5	0.5	1	5
S11	Application of Social Network Analysis to the Study of Open Source Software [152]	Project organization	2006	1	1	1	0.5	1	1	5.5
S12	Who connects with whom? A social network analysis of an online open source software community [124]	Project organization	2011	1	1	1	1	1	1	6

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S13	Network effects: The influence of structural capital on open source project success [126]	Project organization	2011	1	1	1	1	1	1	6
S14	The Onion has Cancer: Some Social Network Analysis Visualizations of Open Source Project Communication [106]	Communication management	2010	1	1	1	1	1	1	6
S15	Evaluating Longitudinal Success of Open Source Software Projects: A Social Network Perspective [148]	Communication management	2009	1	1	1	0.5	1	1	5.5
S16	Using Social Network Analysis for Software Project Management [53]	Risk management	2009	1	-	-	0.5	0.5	1	3
S17	Relating and Clustering Free/Libre Open Source Software Projects and Developers: A Social Network Perspective [71]	Project organization	2011	1	1	1	1	1	1	6
S18	Application of Social Network Theory to Software Development: The problem of task allocation [2]	Project organization	2005	1	1	1	1	1	1	6
S19	Using Social Network Analysis to Inform Management of Open Source Software Development [84]	Project organization	2015	1	1	1	0.5	1	1	5.5
S20	A structured approach to predicting and managing technical interactions in software development (Sosa 2008)	Communication management	2008	1	1	1	1	1	1	6
S21	All-for-One and One-for-All? A Multi-Level Analysis of Communication Networks and Individual Performance in Geographically Distributed Software Development [48]	Communication management	2012	1	1	1	1	1	1	6
S22	Social Network Analysis on Communications for Knowledge Collaboration in OSS Communities [77]	Knowledge management	2006	1	1	1	1	1	1	6
S23	Defect Prediction Using Social Network Analysis on Issue Repositories [9]	Risk management	2011	1	1	1	1	1	1	6
S24	Mining Email Archives and Simulating the Dynamics of Open-Source Project Developer Networks [155]	Project organization	2008	1	1	1	0.5	1	1	5.5
S25	Evolution of Open Source Communities [145]	Project organization	2006	1	1	1	0.5	1	1	5.5

Study ID	**Paper**	**Research area**	**Year**	**QA1**	**QA2 a**	**QA2 b**	**QA3**	**QA4**	**QA5**	**Score**
S26	Impact of Social Ties on Open Source Project Team Formation [61]	Project organization	2006	1	1	1	1	1	1	6
S27	Measuring Coordination Gaps of Open Source Groups Through Social Networks [49]	Project organization	2009	1	1	1	0.5	1	1	5.5
S28	Open source software development and the small world phenomenon: An empirical investigations of macro level collaboration network properties on project success [125]	Project organization	2007	1	1	1	1	1	1	6
S29	Social dynamics of free and open source team communications [69]	Communication management	2006	1	1	1	0.5	1	1	5.5
S30	Socio-Technical Developer Networks: Should We Trust Our Measurements? [100]	Project organization	2011	1	1	1	1	1	1	6
S31	Understanding a Developer Social Network and its Evolution [67]	Project organization	2011	1	1	1	1	1	1	6
S32	StakeNet: Using Social Networks to Analyse the Stakeholders of Large-Scale Software Projects [92]	Requirement management	2010	1	1	1	1	1	1	6
S33	Gauging Influence in Software Development Teams [98]	Project organization	2015	1	1	1	1	1	1	6
S34	Communication Networks in Geographically Distributed Software Development [22]	Communication management	2008	1	1	1	1	1	1	6
S35	Information and influence in social network of the open source community [23]	Knowledge management	2016	1	1	1	0.5	1	1	5.5
S36	Stochastic actor-oriented modeling for studying homophily and social influence in OSS projects [79]	Project organization	2017	1	1	1	1	1	1	6
S37	Who Can Help to Review This Piece of Code? [80]	Project organization	2016	1	1	1	1	1	1	6
S38	Software teams and their knowledge networks in large-scale software development [127]	Knowledge management	2017	1	1	1	1	1	1	6
S39	Knowledge Flows Within Open Source Software Projects: A Social Network Perspective [81]	Knowledge management	2016	1	1	1	1	1	1	6
S40	From Periphery to Core: A Temporal Analysis of GitHub Contributors ' Collaboration Network [21]	Project organization	2017	1	1	1	1	1	1	6

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S41	Boundary Spanners in Open Source Software Development [123]	Project organization	2017	1	1	1	1	1	1	6
S42	Classifying Developers into Core and Peripheral: An Empirical Study on Count and Network Metrics [75]	Project organization	2017	1	1	1	1	1	1	6
S43	Evolutionary trends of developer coordination: a network approach [76]	Project organization	2017	1	1	1	1	1	1	6
S44	Implications of Alter Project Resources and Participant Roles for Open Source Software Project Commercial Success [43]	Project organization	2017	1	1	1	1	1	1	6
S45	Software Quality and Community Structure in Java Software Networks [33]	Project organization	2017	1	1	1	1	1	1	6
S46	Co-membership, Networks Ties, and OSS Success: An Investigation Controlling for Alternative Mechanisms of Knowledge Flow [154]	Knowledge management	2018	1	1	1	0.5	1	1	5.5
S47	The relation between developers' communication and fix-Inducing changes: An empirical study [7]	Communication management	2018	1	1	1	1	1	1	6
S48	Collaboration in the open-source arena: The WebKit case [131]	Project organization	2014	1	1	1	0.5	1	1	5.5
S49	Evolving Relationships between Social Networks and Stakeholder Involvement in Software Projects [91]	Requirement management	2011	1	1	1	0.5	1	1	5.5
S50	Latent Social Structure in Open Source Projects [11]	Project organization	2008	1	1	1	0.5	1	1	5.5
S51	Mining Email Social Networks [10]	Communication management	2006	1	1	1	0.5	1	1	5.5
S52	Reviewer recommendation for pull-requests in GitHub: What can we learn from code review and bug assignment? [156]	Version and configuration management	2016	1	1	1	1	1	1	6
S53	Structural Changes Associated with the Temporal Dispersion of Teams: Evidence from Open Source Software Projects [31]	Project organization	2014	1	1	1	1	1	1	6

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S54	Using Social Network Analysis for Mining Collaboration Data in a Defect Tracking System for Risk and Vulnerability Analysis [130]	Risk management	2011	1	1	1	0.5	1	1	5.5
S55	Emergence of Developer Teams in the Collaboration Network [19]	Project organization	2013	1	1	1	1	1	1	6
S56	Exploring the Impact of Leadership Competencies on Team Social Capital and Performance in IT Service Team [87]	Project organization	2013	1	1	1	0.5	1	1	5.5
S57	From Developer Networks to Verified Communities: A Fine-Grained Approach [74]	Project organization	2015	1	1	1	1	1	1	6
S58	Studying the Impact of Social Structures on Software Quality [8]	Risk management	2010	1	1	1	1	1	1	6
S59	The Importance of Social Network Structure in the Open Source Software Developer Community [138]	Project organization	2010	1	1	1	0.5	1	1	5.5
S60	Analysis of virtual communities supporting OSS projects using social network analysis [135]	Knowledge management	2010	1	1	1	0.5	1	1	5.5
S61	Human agency, social networks, and FOSS project success [141]	Project organization	2012	1	1	1	1	1	1	6
S62	Identifying knowledge brokers that yield software engineering knowledge in OSS projects [129]	Knowledge management	2006	1	1	1	1	1	1	6
S63	Network ties and the success of open source software development [110]	Knowledge management	2013	1	1	1	1	1	1	6
S64	Social networks and coordination performance of distributed software development teams [68]	Project organization	2009	1	1	1	0.5	1	1	5.5
S65	Task and Time Aware Community Detection in Dynamically Evolving Social Networks [63]	Project organization	2013	1	1	1	0.5	1	1	5.5
S66	Understanding the attitudes, knowledge sharing behaviors and task performance of core developers: A longitudinal study [90]	Knowledge management	2014	1	1	1	1	1	1	6

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S67	Overcoming Knowledge Management Challenges During ERP Implementation: The Need to Integrate and Share Different Types of Knowledge [109]	Knowledge management	2007	1	-	-	0.5	1	1	3.5
S68	Measuring social networks when forming information system project teams [85]	Project organization	2017	1	1	1	1	1	1	6
S69	The social structure of Free and Open Source software development [39]	Communication management	2005	1	1	1	0.5	1	1	5.5
S70	Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems [60]	Project organization	2006	1	1	1	1	1	1	6
S71	Hierarchy and Centralization in Free and Open Source Software Team Communications [40]	Project organization	2006	1	1	1	1	1	1	6
S72	Socialization in an Open Source Software community: A socio-technical analysis. [47]	Project organization	2005	1	1	1	1	1	1	6
S73	Returns from social capital in open source software networks [99]	Knowledge management	2009	1	1	1	0.5	0.5	1	5
S74	Joint Effect of Team Structure and Software Architecture in Open Source Software Development [105]	Project organization	2013	1	1	1	1	1	1	6
S75	Developer Initiation and Social Interactions in OSS: A Case Study of the Apache Software Foundation [58]	Communication Management	2015	1	1	1	1	1	1	6
S76	Network Analysis of Software Repositories: Identifying Subject Matter Experts [45]	Knowledge management	2013	1	1	1	0.5	1	1	5.5
S77	Mining and Visualizing Developer Networks from Version Control Systems [74]	Project organization	2011	1	1	1	1	1	1	6
S78	Exploitation and Exploration Networks in Open Source Software Development: An Artifact-Level Analysis [133]	Project organization	2015	1	1	1	1	1	1	6
S79	Predicting Failures with Developer Networks and Social Network Analysis [101]	Risk management	2008	1	1	1	1	1	1	6
S80	Categorizing bugs with social networks: a case study on four open source software communities [157]	Version and configuration management	2013	1	1	1	1	1	1	6

Study ID	Paper	Research area	Year	QA1	QA2 a	QA2 b	QA3	QA4	QA5	Score
S81	Virtual communities as a resource for the development of OSS projects: the case of Linux ports to embedded [136]	Project organization	2009	1	1	1	1	1	1	6
S82	The Role of Domain Knowledge and Cross-Functional Communication in Socio-Technical Coordination [42]	Knowledge management	2013	1	1	1	1	1	1	6
S83	Analysis of Coordination Between Developers and Users in the Apache Community [119]	Project organization	2008	1	1	1	0.5	1	1	5.5
S84	Using Structural Holes Metrics from Communication Networks to Predict Change Dependencies [5]	Communication management	2014	1	1	1	1	1	1	6
S85	Lessons learned from applying social network analysis on an industrial Free/Libre/Open Source Software ecosystem [132]	Project organization	2015	1	1	1	1	1	1	6
S86	Mining Social Networks of Open Source CVE Coordination [117]	Project organization	2017	1	1	1	1	1	1	6

References

[1]Altmann, J. and Pomberger, G. 1999. Kooperative Softwareentwicklung: Konzepte, Modell und Werkzeuge. In *Electronic Business Engineering*, Nüttgens, Scheer, Prof. Dr. Dr. h.c. August-Wilhelm (Hg.), Ed., 643–664.

[2]Amrit, C. 2005. Application of social network theory to software development: the problem of task allocation. In *Computer Supported Activity Coordination (CSAC) // Computer supported activity coordination*, J. Cordeiro, Ed. INSTICC Press, Setúbal.

[3]Amrit, C., Daneva, M., and Damian, D. 2014. Human factors in software development. on its underlying theories and the value of learning from related disciplines. a guest editorial introduction to the special issue. *Information and Software Technology* 56, 12, 1537–1542.

[4]Amrit, C. and van Hillegersberg, J. 2007. Mapping social network to software architecture to detect structure clashes in agile software development. In *European Conference on Information Systems (ECIS) 2007*, Hubert Österle, Joachim Schelp and Robert Winter, Eds. University of St. Gallen, Switzerland.

[5]Baloian, N., Burstein, F., Ogata, H., Santoro, F., Zurita, G., Scaliante Wiese, I., Takashi Kuroda, R., Nassif Roma Junior, D., Ré, R., Ansaldi Oliva, G., and Aurelio Gerosa, M., Eds. 2014. *Using Structural Holes Metrics from Communication Networks to Predict Change Dependencies. Collaboration and Technology*. Springer International Publishing.

[6]Barabási, A.-L. and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286, 5439, 509–512.

[7]Bernardi, M. L., Canfora, G., Di Lucca, G. A., Di Penta, M., and Distante, D. 2018. The relation between developers' communication and fix-Inducing changes. An empirical study. *Journal of Systems and Software* 140, 111–125.

[8]Bettenburg, N. and Hassan, A. E. 2013. Studying the impact of social interactions on software quality. *Empirical Software Engineering* 18, 2, 375–431.

[9]Biçer, S., Bener, A. B., Çağlayan, B., and Raffo, D. 2011. Defect prediction using social network analysis on issue repositories. In *International Conference on Software and Systems Process (ICSSP)*.

[10]Bird, C., Gourley, A., Devanbu, P., Gertz, M., and Swaminathan, A. 2006. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*. ACM, New York, NY, 137. DOI=10.1145/1137983.1138016.

[11]Bird, C., Pattison, D., D'Souza, R., Filkov, V., Devanbu, P., and Harrold, M. J. 2008. Latent social structure in open source projects. In *International Symposium on Foundations of software engineering (FSE)*. ACM, 24. DOI=10.1145/1453101.1453107.

[12]Boehm, B. W. 1988. A spiral model of software development and enhancement. *IEEE Computer* 21, 5, 61–72.

[13]Boeker, M., Vach, W., and Motschall, E. 2013. Google Scholar as replacement for systematic literature searches. Good relative recall and precision are not enough. *BMC medical research methodology* 13, 131.

[14]Borgatti, S. P. and Foster, P. C. 2003. The network paradigm in organizational research. A review and typology. *Journal of Management* 29, 6, 991–1013.

[15]Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software* 80, 4, 571–583.

[16]Broy, M., Jarke, M., Nagl, M., and Rombach, D. 2006. Manifest. Strategische Bedeutung des Software Engineering in Deutschland. *Informatik Spektrum* 29, 3, 210–221.

[17]Broy, M. and Kuhrmann, M. 2013. *Projektorganisation und Management im Software Engineering*. Springer Vieweg, Berlin, Heidelberg.

[18]Burt, R. S. 2009. *Structural holes: the social structure of competition*. Harvard university press, Cambridge, Mass.

[19]Caglayan, B., Bener, A. B., and Miranskyy, A. 2013. Emergence of developer teams in the collaboration network. In 2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE). 25 May 2013, San Francisco, CA, USA ; [part of the 35th International Conference on Software

Engineering (ICSE)]. IEEE, Piscataway, NJ, 33–40. DOI=10.1109/CHASE.2013.6614729.

[20]Çaglayan, B. and Bener, A. B. 2016. Effect of developer collaboration activity on software quality in two large scale projects. *Journal of Systems and Software* 118, C, 288–296.

[21]Camarinha-Matos, L. M., Afsarmanesh, H., Fornasiero, R., El Asri, I., Kerzazi, N., Benhiba, L., and Janati, M., Eds. 2017. *From Periphery to Core: A Temporal Analysis of GitHub Contributors' Collaboration Network. Collaboration in a Data-Rich World*. Springer International Publishing.

[22]Cataldo, M. and Herbsleb, J. D. 2008. Communication networks in geographically distributed software development. In *the ACM 2008 conference*, 579. DOI=10.1145/1460563.1460654.

[23]Celińska, D., Ed. 2016. Information and influence in social network of the open source community.

[24]Centre for Reviews and Dissemination. 2009. *CRD's guidance for undertaking reviews in healthcare*. Systematic reviews. York Publ. Services, York.

[25]Charette, R. N. 2005. Why software fails. *IEEE Spectrum* 42, 9, 42–49.

[26]Chen, X. and Pan, Y.-h. 2014. The study of open source software collaborative user model based on social network and tag similarity. Journal of Electronic Commerce Research 15, 1.

[27]Chesbrough, H. W. and Appleyard, M. M. 2007. Open innovation and strategy. *California Management Review* 50, 1, 57–76.

[28]Chow, T. and Cao, D.-B. 2008. A survey study of critical success factors in agile software projects. *Journal of Systems and Software* 81, 6, 961–971.

[29]Clancy, T. 2014. *The Standish Group Report. CHAOS Report*. https://www.projectsmart.co.uk/white-papers/chaos-report.pdf. Accessed 14 August 2016.

[30]CMMI. 2006. *CMMI for Development, Version 1.2*. www.sei.cmu.edu/reports/06tr008.pdf. Accessed 21 November 2015.

[31]Colazo, J. 2014. Structural changes associated with the temporal dispersion of teams: evidence from open source software projects. In *IEEE 8th International Symposium on Service-Oriented System Engineering (SOSE), 2014. 7 - 11 April 2014, Oxford, United Kingdom ; [including workshop/simposium papers]*. IEEE, Piscataway, NJ, 300–309. DOI=10.1109/HICSS.2014.45.

[32]Comino, S. and Manenti, F. M. 2003. Open source vs closed source software: Public policies in the software market. *Industrial Organization* 306001.

[33]Concas, G., Marchesi, M., Monni, C., Orrù, M., and Tonelli, R. 2017. Software Quality and Community Structure in Java Software Networks. *Int. J. Soft. Eng. Knowl. Eng.* 27, 07, 1063–1096.

[34]Conway, M. E. 1968. How do committees invent. *Datamation* 14, 4, 28–31.

[35]Cooper, H. M. 2003. *Synthesizing research. A guide for literature reviews*. Applied social research methods series 2. Sage, California.

[36]Core. 2018. *CORE Rankings Portal - Computing Research & Education*. http://www.core.edu.au/conference-portal. Accessed 29 August 2018.

[37]Cross, R., Borgatti, S. P., and Parker, A. 2002. Making invisible work visible. using social network analysis to support strategic collaboration. *California Management Review* 44, 2, 25–46.

[38]Cross, R. L. and Parker, A. 2004. The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations. Harvard Business Review Press.

[39]Crowston, K. and Howison, J. 2005. The social structure of free and open source software development. *First Monday* 10, 2.

[40]Crowston, K. and Howison, J. 2006. Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy* 18, 4, 65–85.

[41]Cugola, G. and Ghezzi, C. 1998. Software processes: a retrospective and a path to the future. *Software Process: Improvement and Practice* 4, 3, 101–123.

[42]Damian, D., Helms, R., Kwan, I., Marczak, S., and Koelewijn, B. 2013. *The role of domain knowledge and cross-functional communication in socio-technical coordination*. IEEE Press. http://dl.acm.org/

ft_gateway.cfm?id=2486847&type=pdf.
[43]Daniel, S. and Stewart, K. 2017. Implications of Alter Project Resources and Participant Roles for Open Source Software Project Commercial Success. *ICIS 2017 Proceedings*.
[44]DeMarco, T. and Lister, T. 2013. *Peopleware. productive projects and teams*. Addison-Wesley, Upper Saddle River NJ u.a.
[45]Dittrich, A., Gunes, M. H., and Dascalu, S. 2013. Network analysis of software repositories: identifying subject matter experts. In *Complex Networks*, R. Menezes, A. Evsukoff and M. C. González, Eds. Studies in Computational Intelligence. Springer Berlin Heidelberg, 187–198.
[46]dos Santos, T. A., Araujo, R. M. de, and Magdaleno, A. M. 2010. Identifying collaboration patterns in software development social networks. *Journal of Computer Science*.
[47]Ducheneaut, N. 2005. Socialization in an Open Source Software Community. A Socio-Technical Analysis. *Computer Supported Cooperative Work (CSCW)* 14, 4, 323–368.
[48]Ehrlich, K. and Cataldo, M. 2012. All-for-one and one-for-all?: a multi-level analysis of communication patterns and individual performance in geographically distributed software development. In *Computer Supported Cooperative Work (CSCW)*.
[49]Feczak, S. and Hossain, L. 2009. Measuring coordination gaps of open source groups through social networks. In *International Conference on Enterprise (ICEIS)*.
[50]Fershtman, C. and Gandal, N. 2011. Direct and indirect knowledge spillovers: the "social network" of open-source projects. *The RAND Journal of Economics* 42, 1, 70–91.
[51]Feyhl, A. W. 2004. Management und Controlling von Softwareprojekten: Software wirtschaftlich auswählen, entwickeln, einsetzen und nutzen. Gabler, Wiesbaden.
[52]Fischbach, K., Gloor, P. A., and Schoder, D. 2009. Analysis of informal communication networks – a case study. *Bus. Inf. Syst. Eng.* 1, 2, 140–149.
[53]Fitsilis, P., Gerogiannis, V., Anthopoulos, L., and Kameas, A. 2009. Using social network analysis for software project management. In *Current Trends in Information Technology (CTIT). DOI*=10.1109/CTIT.2009.5423144.
[54]Freeman, L. C. 1978. Centrality in social networks conceptual clarification. *Social Networks* 1, 3, 215–239.
[55]Frischbier, S., Gesmann, M., Mayer, D., Roth, A., and Webel, C. 2012. Emergence as competitive advantage - engineering tomorrow's enterprise software systems. In *International Conference on Enterprise Information Systems (ICEIS)*.
[56]Gebhardt, A. 2003. *Rapid Prototyping*. Carl Hanser Verlag GmbH & Co. KG, München.
[57]Gerber, A., Molefe, O., and van der Merwe, A. 2010. Documenting open source migration processes for re-use. In *Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*, 75–85. DOI=10.1145/1899503.1899512.
[58]Gharehyazie, M., Posnett, D., Vasilescu, B., and Filkov, V. 2015. Developer initiation and social interactions in OSS. A case study of the Apache Software Foundation. *Empir Software Eng* 20, 5, 1318–1353.
[59]Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology* 78, 6, 1360–1380.
[60]Grewal, R., Lilien, G. L., and Mallapragada, G. 2006. Location, Location, Location. How Network Embeddedness Affects Project Success in Open Source Systems. *Management Science* 52, 7, 1043–1056.
[61]Hahn, J., Moon, J. Y., and Zhang, C. 2006. Impact of social ties on open source project team formation. In *Open source systems*. Springer, 307–317.
[62]Häsel, M. 2007. Softwareentwicklung in internetbasierten Gründungsunternehmen – Anforderungen an Produkte, Prozesse und Personal. *Informatik Spektrum* 30, 5, 327–335.
[63]Hecking, T., Göhnert, T., Zeini, S., and Hoppe, U. 2013. Task and Time Aware Community Detection in Dynamically Evolving Social Networks. *Procedia Computer Science* 18, 2066–2075.
[64]Henrich, A. 2002. *Management von Softwareprojekten*. Oldenbourg, München.
[65]Herbsleb, J. D. and Moitra, D. 2001. Global software development. *IEEE Softw.* 18, 2, 16–20.
[66]Holmstrom, H., Conchuir, E., Agerfalk, P., and Fitzgerald, B. 2006. Global software development challenges: a

case study on temporal, geographical and socio-Cultural distance. In *International Conference on Global Software Engineering (ICGSE)*, 3–11. DOI=10.1109/ICGSE.2006.261210.

[67]Hong, Q., Kim, S., Cheung, S. C., and Bird, C. 2011. Understanding a developer social network and its evolution. In *International Conference Software Maintenance (ICSM).*

[68]Hossain, L. and Zhu, D. 2009. Social networks and coordination performance of distributed software development teams. *The Journal of High Technology Management Research* 20, 1, 52–61.

[69]Howison, J., Inoue, K., and Crowston, K. 2006. Social dynamics of free and open source team communications. In *Open source systems*. Springer, 319–330.

[70]Hu, D. and Zhao, J. L. 2009. Discovering determinants of project participation in an open source social network. *ICIS 2009 Proceedings*, 16.

[71]Huang, K.-Y. and Choi, N. 2011. Relating and clustering free/libre open source software projects and developers: a social network perspective. In *Hawaii International Conference on System Sciences (HICSS).*

[72]InCites. 2018. *InCites*. http://webofknowledge.com/JCR. Accessed 29 August 2018.

[73]ISO/ IEC 12207:2008. *Information technology – Software lifecycle processes*. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43447.

[74]Jermakovics, A., Sillitti, A., and Succi, G. 2011. Mining and visualizing developer networks from version control systems. In *International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE).*

[75]Joblin, M., Apel, S., Hunsen, C., and Mauerer, W. 2017. *Classifying developers into core and peripheral. An empirical study on count and network metrics*. IEEE Press. http://dl.acm.org/ft_gateway.cfm?id=3097389&type=pdf.

[76]Joblin, M., Apel, S., and Mauerer, W. 2017. Evolutionary trends of developer coordination. A network approach. *Empirical Software Engineering* 22, 4, 2050–2094.

[77]Kakimoto, T., Kamei, Y., Ohira, M., and Matsumoto, K. 2006. Social network analysis on communications for knowledge collaboration in oss communities. In *International Workshop on Supporting Knowledge Collaboration in Software Development (KCSD).*

[78]Kappelhoff, P. 1987. Cliquenanalyse. Die Bestimmung von intern verbundenen Teilgruppen in Netzwerken. *Techniken der empirischen Sozialforschung–Methoden der Netzwerkanalyse. München: Pieper*, 39–63.

[79]Kavaler, D. and Filkov, V. 2017. Stochastic actor-oriented modeling for studying homophily and social influence in OSS projects. *Empir Software Eng* 286, 5439, 509.

[80]Kerzazi, N. and Asri, I. E. 2016. Who can help to review this piece of code? In . Springer, Cham, 289–301. DOI=10.1007/978-3-319-45390-3_25.

[81]Kerzazi, N. and Asri, I. E. 2017. *Knowledge flows within open source software projects: a social network perspective*. Springer, Singapore. https://link.springer.com/content/pdf/10.1007/978-981-10-1627-1_19.pdf.

[82]Kitchenham, B. and Charters, S. 2007. Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report. EBSE.*

[83]Königs, H.-P. 2009. IT-Risiko-Management mit System. Von den Grundlagen bis zur Realisierung ; ein praxisorientierter Leitfaden. Edition kes. Vieweg + Teubner, Wiesbaden.

[84]Kumar, S. 2015. Using social network analysis to inform management of open source software development. In *Hawaii International Conference on System 2015*, 5154–5163.

[85]Latorre, R. and Suárez, J. 2017. Measuring social networks when forming information system project teams. *Journal of Systems and Software* 134, 304–323.

[86]Laurent, A. M. S. 2004. *Understanding Open Source and Free Software Licensing*. EBL-Schweitzer. O'Reilly Media, Sebastopol.

[87]Lee, J., Lee, H., and Park, J.-g. 2013. Exploring the Impact of Leadership Competencies on Team Social Capital and Performance in IT Service Team. In *2013 46th Hawaii International Conference on System Sciences (HICSS 2013). Wailea, [Maui], Hawaii, USA, 7 - 10 January 2013 ; [proceedings]*. IEEE, Piscataway, NJ,

4344–4353. DOI=10.1109/HICSS.2013.224.

[88]Leffingwell, D. and Widrig, D. 2000. *Managing Software Requirements: A Unified Approach*. Addison-Wesley Professional; Addison-Wesley, Reading, MA.

[89]2010. Leveraging Applications of Formal Methods, Verification and Validation. Springer.

[90]Licorish, S. A. and MacDonell, S. G. 2014. Understanding the attitudes, knowledge sharing behaviors and task performance of core developers. A longitudinal study. *Information and Software Technology* 56, 12, 1578–1596.

[91]Lim, S. L. and Bentley, P. J. 2011. Evolving relationships between social networks and stakeholder involvement in software projects. In *ACM 13th Genetic and Evolutionary Computation (GECCO).*

[92]Lim, S. L., Quercia, D., and Finkelstein, A. 2010. StakeNet: using social networks to analyse the stakeholders of large-scale software projects. In *ICSE '10 Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering*.

[93]Long, Y. and Siau, K. 2006. Social network dynamics for open source software projects. In *Americas Conference on Information Systems (AMCIS).*

[94]Long, Y. and Siau, K. 2008. Impacts of social network structure on knowledge sharing in open source software development teams. In *Americas Conference on Information Systems (AMCIS).*

[95]Madey, G., Freeh, V., and Tynan, R. 2002. The open source software development phenomenon: An analysis based on social network theory. In *Americas Conference on Information Systems (AMCIS).*

[96]Majchrzak, T. A. 2012. Organizational aspects: the IHK project. In *Improving Software Testing. Technical and Organizational Developments*, T. A. Majchrzak, Ed. SpringerBriefs in Information Systems. Springer, Heidelberg, 95–109. DOI=10.1007/978-3-642-27464-0_5.

[97]Manteli, C., van Vliet, H., and van den Hooff, B. 2012. Adopting a social network perspective in global software development. In *Global Software Engineering (ICGSE).*

[98]Marshall, A. and Gamble, R. 2015. Gauging influence in software development teams. In *2015 IEEE Frontiers in Education Conference (FIE). 21 - 24 Oct. 2015, El Paso, Texas : proceedings*. IEEE, Piscataway, NJ, 1–8. DOI=10.1109/FIE.2015.7344106.

[99]Méndez-Durón, R. and García, C. E. 2009. Returns from social capital in open source software networks. *Journal of Evolutionary Economics* 19, 2, 277–295.

[100]Meneely, A. and Williams, L. 2011. Socio-technical developer networks: should we trust our measurements? In *International Conference on Software Engineering (ICSE).*

[101]Meneely, A., Williams, L., Snipes, W., and Osborne, J. 2008. *Predicting failures with developer networks and social network analysis*. ACM. http://dl.acm.org/ft_gateway.cfm?id=1453106&type=pdf.

[102]Moll, K.-R., Broy, M., Pizka, M., Seifert, T., Bergner, K., and Rausch, A. 2004. Erfolgreiches Management von Software-Projekten. *Informatik-Spektrum* 27, 5, 419–432.

[103]Müller, C. 2007. Analyse sozialer Netzwerke und Social Software - Grundlagen und Anwendungsbeispiele. Gito, Berlin.

[104]Nahapiet, J. and Ghoshal, S. 1998. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review* 23, 2, 242–266.

[105]Nan, N. and Kumar, S. 2013. Joint Effect of Team Structure and Software Architecture in Open Source Software Development. *IEEE Trans. Eng. Manage.* 60, 3, 592–603.

[106]Oezbek, C., Prechelt, L., and Thiel, F. 2010. The onion has cancer: some social network analysis visualizations of open source project communication. In *International Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development (FLOSS).*

[107]Oliveira, E., Conte, T., Cristo, M., and Valentim, N. 2019. Influence Factors in Software Productivity — A Tertiary Literature Review. *Int. J. Soft. Eng. Knowl. Eng.* 28, 11n12, 1795–1810.

[108]Ouriques, R. A. B., Wnuk, K., Gorschek, T., and Svensson, R. B. 2019. Knowledge Management Strategies and Processes in Agile Software Development. A Systematic Literature Review. *Int. J. Soft. Eng. Knowl. Eng.* 29, 03, 345–380.

[109]Pan, S. L., Newell, S., Huang, J., and Galliers, R. D. 2007. Overcoming knowledge management challenges during ERP implementation. The need to integrate and share different types of knowledge. *Journal of the American Society for Information Science and Technology* 58, 3, 404–419.

[110]Peng, G., Wan, Y., and Woodlock, P. 2013. Network ties and the success of open source software development. *The Journal of Strategic Information Systems* 22, 4, 269–281.

[111]Perry, W. E. 2006. *Effective methods for software testing*. Safari Books Online. Wiley, Indianapolis, Ind.

[112]Portillo-Rodríguez, J., Vizcaíno, A., Piattini, M., and Beecham, S. 2012. Tools used in Global Software Engineering. A systematic mapping review. *Information and Software Technology* 54, 7, 663–685.

[113]Pryke, S. and Smyth, H. 2006. *The management of complex projects. a relationship approach*. Blackwell Pub, Oxford.

[114]Rosenstreich, D. and Wooliscroft, B. 2009. Measuring the impact of accounting journals using Google Scholar and the g-index. *The British Accounting Review* 41, 4, 227–239.

[115]Royce, W. W., Ed. 1970. Managing the development of large software systems 26. Los Angeles.

[116]Runeson, P., Rainer, A., and Regnell, B. 2012. *Case study research in software engineering*. John Wiley & Sons.

[117]Ruohonen, J., Hyrynsalmi, S., Rauti, S., and Leppänen, V. 2017. *Mining social networks of open source CVE coordination*. ACM. http://dl.acm.org/ft_gateway.cfm?id=3143458&type=pdf.

[118]Ruparelia, N. B. 2010. Software development lifecycle models. *SIGSOFT Softw. Eng. Notes* 35, 3, 8.

[119]Russo, B., Damiani, E., Hissam, S., Lundell, B., Succi, G., Kamei, Y., Matsumoto, S., Maeshima, H., Onishi, Y., Ohira, M., and Matsumoto, K.-i., Eds. 2008. *Analysis of Coordination Between Developers and Users in the Apache Community. Open Source Development, Communities and Quality*. Springer US.

[120]Schenk, M. 1995. Soziale Netzwerke und Massenmedien. Untersuchungen zum Einfluß der persönlichen Kommunikation. Mohr, Tübingen.

[121]Scott, J. 2013. *Social network analysis*. Sage, London.

[122]Sharma, M. 2014. Knowledge management: principle and strategy. *Sai Om Journal of Commerce & Management* 1, 10, 7–15.

[123]Sharma, P. N., Savarimuthu, B. T. R., and Stanger, N. 2017. Boundary Spanners in Open Source Software Development. A Study of Python Email Archives. In *24th Asia-Pacific Software Engineering Conference (APSEC 2017). 4-8 December 2017, Nanjing, Jiangsu, China : proceedings*. IEEE, Piscataway, NJ, 308–317. DOI=10.1109/APSEC.2017.37.

[124]Shen, C. and Monge, P. 2011. Who connects with whom? A social network analysis of an online open source software community. *First Monday* 16, 6.

[125]Singh, P. 2007. Open source software development and the small world phenomenon: An empirical investigations of macro level colaboration network properties on project success. *ICIS 2007 Proceedings*, 18.

[126]Singh, P. V., Tan, Y., and Mookerjee, V. 2011. Network effects: the influence of structural capital on open source project success. *MIS Quarterly* 35, 4, 813–830.

[127]Šmite, D., Moe, N. B., Šāblis, A., and Wohlin, C. 2017. Software teams and their knowledge networks in large-scale software development. *Information and Software Technology* 86, 71–86.

[128]Sosa, M. E. 2008. A structured approach to predicting and managing technical interactions in software development. *Research in Engineering Design* 19, 1, 47–70.

[129]Sowe, S., Stamelos, I., and Lefteris, A. 2006. Identifying knowledge brokers that yield software engineering knowledge in OSS projects. *Information and Software Technology* 48, 11, 1025–1033.

[130]Sureka, A., Goyal, A., and Rastogi, A. 2011. Using social network analysis for mining collaboration data in a defect tracking system for risk and vulnerability Analysis. In *India Software Engineering Conference (ISEC). DOI*=10.1145/1953355.1953381.

[131]Teixeira, J. and Lin, T. 2014. *Collaboration in the open-source arena. The webkit case*. ACM. http://dl.acm.org/ft_gateway.cfm?id=2600009&type=pdf.

[132]Teixeira, J., Robles, G., and González-Barahona, J. M. 2015. Lessons learned from applying social network analysis on an industrial Free/Libre/Open Source Software ecosystem. *Journal of Internet Services and Applications* 6, 1, 14.

[133]Temizkan, O. and Kumar, R. L. 2015. Exploitation and exploration networks in open source software development. an artifact-level analysis. *Journal of Management Information Systems* 32, 1, 116–150.

[134]Tichy, N. M., Tushman, M. L., and Fombrun, C. 1979. Social network analysis for organizations. *Academy of Management Review* 4, 4, 507–519.

[135]Toral, S. L., Martínez-Torres, M. R., and Barrero, F. 2010. Analysis of virtual communities supporting OSS projects using social network analysis. *Information and Software Technology* 52, 3, 296–303.

[136]Toral, S. L., Martínez-Torres, M. R., and Barrero, F. J. 2009. Virtual communities as a resource for the development of OSS projects. The case of Linux ports to embedded processors. *Behaviour & Information Technology* 28, 5, 405–419.

[137]Uzzi, B. 1997. Social structure and competition in interfirm networks. the paradox of embeddedness. *Administrative Science Quarterly* 42, 1, 35.

[138]van Antwerp, M. and Madey, G. 2010. The importance of social network structure in the open source software developer community. In *Hawaii International Conference on System 2010*.

[139]Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. Reconstructing the giant: on the importance of rigour in documenting the literature search process. In *European Conference on Information Systems (ECIS)*.

[140]Wagstrom, P., Herbsleb, J., and Carley, K. 2005. A social network approach to free/open source software simulation. In *International Conference on Open Source Systems (OSS)*.

[141]Wang, J., Hu, M. Y., and Shanker, M. 2012. Human agency, social networks, and FOSS project success. *Journal of Business Research* 65, 7, 977–984.

[142]Wasserman, S. and Faust, K. 1994. *Social network analysis. Methods and applications*. Structural analysis in the social sciences 8. Cambridge University Press, Cambridge, New York.

[143]Watts, D. J. 2004. The "new" science of networks. *Annual Review of Sociology* 30, 243–270.

[144]Webster, J. and Watson, R. T. 2002. Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly* 26, 2, xiii–xxiii.

[145]Weiss, M., Moroiu, G., and Zhao, P. 2006. Evolution of open source communities. In *Open source systems*. Springer, 21–32.

[146]Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. 2012. *Experimentation in software engineering*. Springer, Berlin.

[147]Wohlin, C., Šmite, D., and Moe, N. B. 2015. A general theory of software engineering: balancing human, social and organizational capitals. *Journal of Systems and Software* 109, 229–242.

[148]Wu, J. and Goh, K. Y. 2009. Evaluating longitudinal success of open source software projects: A social network perspective. In *Hawaii International Conference on System Sciences (HICSS)*.

[149]Wu, J., Goh, K.-Y., and Tang, Q. 2007. Investigating success of open source software projects: a social network perspective. In *International Conference on Information Systems (ICIS)*, 105.

[150]Wu, J. and Tang, Q. 2007. Analysis of survival of open source projects: a social network perspective. In *Australian Conference of Information Systems (ACIS)*.

[151]Xia, H., Dawande, M., and Mookerjee, V. 2016. Optimal coordination in distributed software development. *Production and Operations Management* 25, 1, 56–76.

[152]Xu, J., Christley, S., and Madey, G. 2006. Application of social network analysis to the study of open source software. In *The economics of open source software development*, J. Bitzer and P. J. H. Schröder, Eds. Elsevier, Amsterdam, 247–269. DOI=10.1016/B978-044452769-1/50012-3.

[153]Yilmaz, M., O'Connor, R. V., and Clarke, P. 2016. Effective Social Productivity Measurements during Software Development — An Empirical Study. *Int. J. Soft. Eng. Knowl. Eng.* 26, 03, 457–490.

[154]Yu, F. and Peng, Y. 2018. Co-membership, Networks Ties, and OSS Success: An Investigation Controlling for Alternative Mechanisms for Knowledge Flow. *Hawaii International Conference on System Sciences 2018 (HICSS-51)*.

[155]Yu, L., Ramaswamy, S., and Zhang, C. 2008. Mining email archives and simulating the dynamics of open-source project developer networks. In *International Workshop on Enterprise and Organizational Modeling and Simulation (EOMAS)*.

[156]Yu, Y., Wang, H., Yin, G., and Wang, T. 2016. Reviewer recommendation for pull-requests in GitHub. What can we learn from code review and bug assignment? *Information and Software Technology* 74, 204–218.

[157]Zanetti, M. S., Scholtes, I., Tessone, C. J., and Schweitzer, F. 2013. Categorizing bugs with social networks: a case study on four open source software communities. In *International Conference on Software Engineering (ICSE)*.

Paper II

Bibliographic data	Zylka, M. P. & Fischbach, K. (2017). *Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda.* In: ACM SIGMIS Database, Vol.48, No. 2, pp. 52-78, https://doi.org/10.1145/3084179.3084185
Status of publication	Published as completed research paper.
Publication type	Journal article
Version & Copyright notice	Accepted version
Previous versions	Zylka, M. P. (2016). *Putting the Consequences of IT Turnover on the Map: A Review and Call for Research.* In: Proceedings of the 2016 ACM SIGMIS Conference on Computers and People Research (SIGMIS-CPR '16), pp. 87–95, Best Paper Nominee. ACM, New York, NY, USA, https://doi.org/10.1145/2890602.2890618

Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda

Among the many aspects of IT personnel studied by the information systems (IS) research community, individual voluntary IT turnover is one of the most-examined phenomena. However, research into this phenomenon concentrates mainly on antecedents and cognitive precursors such as turnover intention. Antecedents are essential to understanding the turnover behavior of IT personnel, but they do not represent the complete IT turnover research. The consequences of individual voluntary IT turnover behavior, an important topic, have faded from the spotlight of the IS community. We investigate this by conducting a multidisciplinary scoping literature review of individual voluntary IT turnover behavior, with the focus on the consequences. The purpose of this review is to determine what is known about voluntary IT turnover behavior consequences and what research gaps exist in this research context. To make our review as rigorous as possible, we followed a systematic review approach, accompanied with transparent reporting of all steps of the review process. Our search strategy yielded 153 IT turnover studies, 14 of which consider IT turnover behavior consequences, concentrated primarily on IT project management. Drawing on the scoping review, our study also specifies a research agenda for future IT turnover behavior consequences research by highlighting knowledge gaps for potentially fruitful research directions.

*Keywords***:** IT turnover; individual voluntary turnover; consequences; effects; scoping review; literature review

Introduction

In recent years, the employment growth rates of IT-related occupations have been consistently higher than those for total employment, reflecting the increasing relevance of IT personnel (see Figure 1). In addition, the technological convergence – a process by which the ICT sectors may be converging towards a unified market (Borés et al., 2003; García-Murillo and MacInnes, 2003; Spiegel et al., 2014) – intensifies the importance of and demand for IT personnel in this unified market. Hence, the competition for and retention of highly skilled IT personnel is critical to organizations. There may be serious consequences if retention strategies do not work and highly skilled IT employees voluntarily leave their employers.

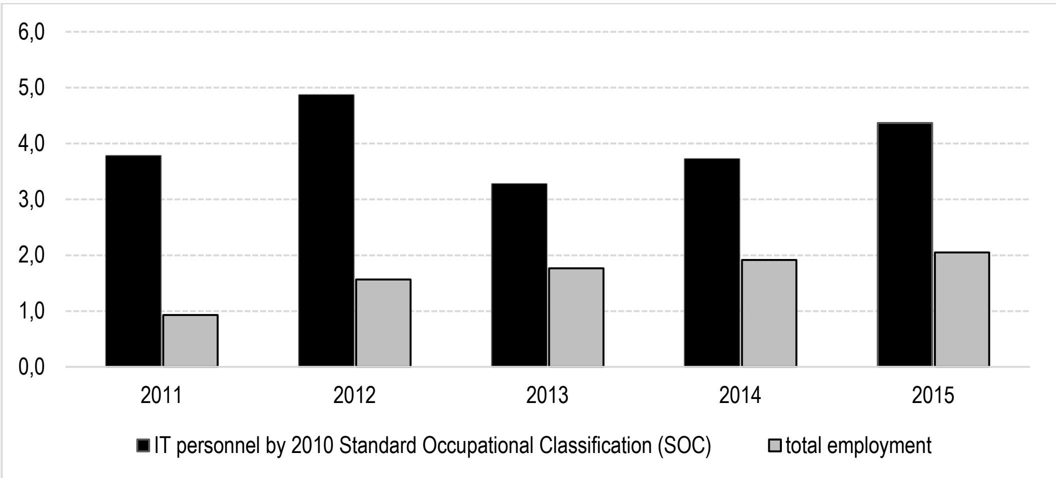


Figure 1. IT personnel employment versus total employment in the USA by growth rate and year (Based on U.S. Bureau of Labor Statistics data, 2016)

Research in IS-related disciplines such as management and applied psychology shows that employee turnover behavior leads to poor job attitudes, high training and recruitment costs, and operational disruption, and diminishes human and social capital (Dess and Shaw, 2001; Mobley, 1977; Pennings et al., 1998; Price, 1977; Staw, 1980). All of these factors translate into negative consequences for individuals and organizations, and especially affect performance and profits (Ton and Huckman, 2008). However, employee turnover behavior may also lead to positive consequences for organizations and individuals (Mobley, 1982). Losing a highly skilled employee to a competitor may lead to access to external knowledge for the former employer (Aime et al., 2010; Somaya et al., 2008).

These turnover behavior consequences are relevant for employees in general and thus for IT personnel. But because differences in personalities and skills between IT personnel and non-IT personnel found by researchers (Bartol and Martin, 1982; Wynekoop and Walz, 1998), turnover behavior of IT personnel may result in consequences different than those from non-IT turnover.

In its research on IT turnover, the information systems (IS) research community has focused mostly on the antecedents or cognitive precursors (e.g., turnover intention) of individual voluntary IT turnover behavior, using or extending common models well established in the applied psychology and management literatures (Joseph et al., 2007; Lo, 2015). For instance, Niederman et al. (2007) use the "unfolding model of voluntary turnover" (Lee and Mitchell, 1994) to examine the turnover process of IT personnel. Eckhardt et al. (2016) propose and test a model of turnover intention across IT job types, combining the "Big Five" personality traits and a basic turnover model originally rooted in management and applied psychology studies (cf. Mobley, 1977; Mobley et al., 1979; Mowday et al., 1982; Porter et al., 1976; Steers, 1977) that has proven reliable for Western IT professionals (Lacity et al., 2008). Further, Joseph et al. (2007) develop a contextual model of IT turnover intention among IT professionals based on the process model of turnover by March and Simon (1958); it has, by far, received the most attention from turnover researchers. Other researchers concentrate on the career paths IT employees may follow after a turnover (Joseph et al., 2012; Mourmant et al., 2009). Mourmant et al. (2009, 2010) focus, for example, on IT entrepreneurial turnover as a specific career path after an individual voluntary turnover. Spiegel et al. (2016) show that individual turnover is useful when employees who leave use their social capital accrued with contacts at former employers to develop their start-ups' business models and ultimately make them successful.

Despite the growing body of research on IT turnover, the consequences of IT turnover behavior remain unclear. It seems the research community has insufficiently studied the consequences of turnover of IT personnel.

The objective of this study is to present the current state of the literature regarding individual IT turnover consequences and examine whether and how IT turnover studies have considered the consequences of individual voluntary IT turnover. Our two research questions, then, are:

- What is the current state of research into the consequences of individual voluntary IT turnover behavior?
- What research gaps exist regarding the consequences of individual voluntary IT turnover behavior consequences?

We answer these questions by conducting a multidisciplinary scoping review of IT turnover studies, with the focus on consequences. Further, we provide a taxonomic overview of individual turnover consequences that can be utilized in future research by the IT turnover research community. We identify research gaps and propose a research agenda for studying these consequences in detail.

This study contributes to research and practice in two major ways. First, it contributes to the understanding of the turnover of IT personnel. With an awareness of the antecedents and consequences of IT turnover, top management can weigh its pros and cons and effectively manage human resources. Second, it points out that there is still a need for research into IT turnover consequences and identifies the research gaps.

The remainder of this paper is structured as follows. The next section outlines related work on the consequences of individual voluntary turnover, as well as the skills and mindsets of IT personnel. A debate about the unique skills and mindsets of IT personnel is necessary to emphasize consequences that might appear only after an IT turnover. The subsequent section describes the literature review, followed by a section that presents the research results. The paper concludes with a discussion of the results and implications for future research.

Background

This section begins with a definition of our study's context and scope, after which we present an overview of the consequences of individual voluntary turnover as discussed generally (i.e., independent of specific occupation) in the applied psychology and management literatures. From this overview, we derive a taxonomy that summarizes these consequences. We then provide an overview of the discussion regarding the skills and mindsets of IT personnel, important for understanding how IT personnel differ from other occupational groups and the unique consequences that may occur when an IT employee leaves an organization.

Context and Definition of Turnover

The context of research is important for evaluating the boundaries of empirical and theoretical claims. Further, the context of research frames and defines the target research population. For the purpose of generalizability, researchers need to know their research context so they can clearly define their target population and discuss potential boundary conditions that might apply to their findings (Seddon and

Scheepers, 2012). Hence, context definition leads to corresponding boundary conditions of the knowledge claims.

Figure 2 presents the context of this study. We examine the consequences of individual IT turnover behavior and do not focus on turnover intention. Turnover behavior refers to leaving a job for a similar, alternative job across organizational boundaries (Joseph et al., 2012; Joseph et al., 2015). Hence, we do not focus our study on other job mobility types such as turnaway-within and turnaway-between mobility. Concerning the turnover level, we pay attention to the consequences of an individual leaving a firm, and do not consider collective turnover, as it is inappropriate to generalize individual turnover to collective turnover. Individual turnover differs from collective turnover conceptually and empirically and has different antecedents and consequences (Nyberg and Ployhart, 2013). Hence, we only present consequences that occur after an individual turnover. However, note that several consequences presented in the following sections may also occur in the collective turnover context.

The literature considers four types of turnover: voluntary, involuntary, functional, and dysfunctional. Turnover initiated by forces other than the leaving employee are called involuntary, while turnover initiated by an employee is called voluntary (Bluedorn, 1978). Functional turnover occurs when a low-performing employee leaves the organization; dysfunctional turnover occurs when a high-performing employee leaves the organization (Dalton et al., 1982). In our study, turnover takes on the role of an antecedent (from a model perspective).

Conceptualization of Turnover	Turnover behavior		Turnover intention	
Turnover Level	Individual		Collective	
Turnover Type	Voluntary	Involuntary	Functional	Dysfunctional
Turnover Role	As antecedent		As consequent	

Figure 2. Context of this turnover study (grayed out)

Consequences of Individual Voluntary Turnover

Individual voluntary turnover behavior is often seen as having negative consequences on individuals, and firms. However, in some cases, employee turnover behavior can have positive impact. In the following, we present an overview of both negative and positive consequences of individual voluntary turnover and who is affected by these consequences.

Negative Consequences

Individual turnover consequences are traditionally considered to be negative (Dalton and Todor, 1979). They increase recruitment and training costs (Dess and Shaw, 2001; Mobley, 1982), are operationally disruptive at the organizational level, lead to demoralization at the individual level (Allen et al., 2001; Staw, 1980), and may affect individuals and organizations (Mobley, 1982; Muchinsky and Morrow, 1980; Steers and Mowday, 1981).

Relationships with co-workers can be affected negatively by individual turnover (Dalton et al., 1982; Krackhardt and Porter, 1985; Krackhardt and Porter, 1986; Nyberg and Ployhart, 2013). Demoralization is a negative consequence in this context (Staw, 1980); it can affect both the former and new co-workers. Co-workers who remain behind may question their own motivations for staying (Staw, 1980). They interpret the departure of a former colleague as a rejection of the job and may begin to realize the possibility that better job opportunities exist (Steers and Mowday, 1981). As Steers and Mowday (1981) state: “Those who remain in the organization may have to reconcile their decision to stay in light of evidence from the behavior of another individual that the job may not be all that desirable.” Remaining co-workers may reevaluate their present positions in the organization and ultimately develop more negative job attitudes, which in turn could initiate a search for a more attractive job (Staw, 1980; Steers and Mowday, 1981). In the context of involuntary turnover (e.g. through downsizing), most companies identified decreased morale as the most probable after-effect (Allen et al., 2001).

Finally, individual turnover has negative consequences for the individual who leaves. He or she may reevaluate both the chosen and unchosen jobs following the choice (Steers and Mowday, 1981). Steers and Mowday (1981) state that turnover may have important implications for attitudes toward the job the individual is leaving, as well as the employment he or she is taking. The manifestation of positive or negative feelings in the process of justifying his or her decision depends on whether the individual’s behavior is consistent with his or her attitude regarding the former job (Steers and Mowday, 1981). According to the cognitive dissonance theory (Festinger, 1962), once an employee is well-embedded in the firm, one has to admit to having positive job attitudes and will be less willing to leave the organization (Hom and Griffeth, 1991; Mowday et al., 1982). If a satisfied employee voluntarily leaves – meaning his or her behavior is inconsistent with that attitude – a turnover will create cognitive dissonance in his or her mind (Steers and Mowday, 1981).

Another negative consequence for the individual who has left is the potential loss of social capital (Dess and Shaw, 2001). Voluntary turnover disrupts established social relations (Pennings et al., 1998). When he or she leaves unexpectedly and remaining co-workers have to compensate by, for example, working overtime, it may lead to bad moods and, as a consequence, remaining co-workers breaking off contact with the departed colleague. The individual who leaves may experience anxiety at his or her new job position (Feldman and Brett, 1983). To reduce the anxiety, attempts (e.g., through socialization) should be made to integrate the person into the informal organization (Cable et al., 2013). Sometimes socialization is not that easy, because arriving employees have unrealistic or unmet expectations regarding the new job or enter unfamiliar organizational settings and experience reality shocks (Louis, 1980). However, arriving employees must adapt to the new organizational environment; they may experience uncertainty the first time in the new job, which leads them to engage in an "information search" by increasingly contacting colleagues and supervisors (Ashford and Cummings, 1983). Although the quantity of communication will increase to socialize and train new members, the quality of this communication will decline because lines of communication will be disrupted and new members will be less capable of sending and receiving messages accurately (Bluedorn, 1982; Muller and Price, 1989).

When an employee leaves an organization, it is usually the case that a new replacement employee must be recruited, selected, and hired. This process involves costs to the organization, the amount of which depends on the level and complexity of the job to be filled (Lindbeck and Snower, 2001; Mobley, 1982; Staw, 1980). The more demanding the job, the costlier the recruitment process (Staw, 1980). Once a new employee is found, he or she must be trained to do the tasks of the former employee; the complexity of the job and the skill set of the new employee determine how much the training costs (Staw, 1980). Jobs with complex tasks have higher training costs than jobs with simple tasks. To minimize recruitment and training costs, organizations may promote or reassign workers to the departed employee's position (Muchinsky and Morrow, 1980). Further, turnover of a highly skilled employee increases the risks of leaking knowledge of the former employer to the competitor (Aime et al., 2010; Somaya et al., 2008).

Turnover may also have a negative effect on operations flows. In particular, the turnover of a key employee – especially where functional roles are closely intertwined – may result in operational disruption (Bluedorn, 1982; Dalton and Todor, 1979; Mobley, 1982; Staw, 1980; Ton and Huckman, 2008). The turnover of an employee who had an important job role in terms of coordination and communication may have a larger impact on operational disruption than that of an employee not involved in critical work tasks (Mobley, 1982). The individual turnover may disrupt routines and it may be more difficult to find replacements and integrate them into work groups because individuals are encouraged to stand out and bring unique skills to the job (Hancock et al., 2013). This may be more of an issue in individualistic cultures and less crucial in collective cultures (such as China, and Korea), where it may be simpler to find and integrate replacements into group functioning (Hancock et al., 2013). It may also be the case that individuals departing in more collective cultures may tend to be those with weaker fits to the work group or larger organization (Hancock et al., 2013). In addition, a recent study by Bermiss and Murmann (2015) shows that the turnover of a top executive whose functional role focuses on internal firm processes is more harmful to the organization than losing a top executive whose functional role focuses on managing external exchange relationships. However, voluntary turnover of leading employees leaves subordinates behind with uncertainty coupled with concern about how they will fare without the departing leader and/or with the incoming leader (Shapiro et al., 2016).

Positive Consequences

Individual turnover may have also positive consequences for the individual who leaves, as well as for co-workers and organizations (Mobley, 1982).

An individual turnover may resolve conflicts. If a conflicting supervisor or co-worker leaves an organization, it may well be a happy occasion for the remaining co-workers who have been in conflict with the departing individual. Even if well-liked and/or productive employees leave an organization, the turnover may open positions in an otherwise impenetrable hierarchy (Staw, 1980). Thus, turnover may be the primary creator of promotion opportunities (Dalton and Todor, 1979), contributing to a positive relation between the turnover and individual morale (Staw, 1980). Further, the departing/arriving employee will achieve higher job satisfaction at his or her new job and a higher level of job commitment. Unlike with the earlier mention of cognitive dissonance, consistency in attitude and behavior that results in a turnover (i.e., a dissatisfied employee leaves voluntarily) will not generate cognitive dissonance in the departing employee's mind (Steers and Mowday, 1981). Further, employees who have voluntarily left many jobs would likely experience and explain voluntary turnover differently (more positive) than employees who have had fewer voluntary departures. They see voluntary turnover as a routine career event, whereas others see turnover as unusual and unnecessary (Lee and Mitchell, 1994).

Further, the newly arrived employee increases his or her social capital and experiences socialization through the new employment (Feldman, 1976). The arrival of a new colleague may improve the organizational commitment of co-workers as well (Yang, 2008) because they may adopt the new employee's positive work attitude. The departure of an employee may increase the organizational commitment of co-workers at the former employer, especially in cases where the job level of the departing colleague is higher than those colleagues remaining, who may see the opportunity for job promotion (Mobley, 1982; Staw, 1980).

There is evidence that turnover increases organizational performance (Staw, 1980; Ton and Huckman, 2008). The argument is that voluntary turnover is good for organizational performance if the new employee possesses better job skills or brings new ideas to the unit (Staw, 1980). Moreover, new employees may be more highly motivated than their longer-tenured counterparts and hence more productive (Muchinsky and Morrow, 1980). However, job performance depends on the job role. New employees in high-stress roles tend to have an inverted U-shaped performance curve (Staw, 1980). They begin with low performance, improve as their tenure lengthens, but end their tenure with low job performance. Staw (1980) argues that most jobs have an inverted U-shaped performance curve because performance is typically a joint function of skills and effort. The literature on the role of individual turnover on organizational performance further highlights the ways in which employee mobility introduces valuable human capital (and the employee's embedded know-how and skills) into target organizations (Carnahan et al., 2012; Corredoira and Rosenkopf, 2010) and facilitates expansion of an organization's breadth (Rosenkopf and Almeida, 2003). Further, the new employee transfers social capital to the new employer (i.e., by relations to former clients (Somaya et al., 2008)), and facilitates the flow of knowledge back to source firms (Aime et al., 2010; Madsen et al., 2003) by enhancing social networks between the former and new employer (Corredoira and Rosenkopf, 2010).

Voluntary turnover of employees also resolves certain staffing problems for the former employer. An ineffective employee who voluntarily quits is conveniently eliminated without interpersonal discomfort or increases in unemployment insurance costs (Muchinsky and Morrow, 1980).

Interpersonal conflicts between co-workers, which exist at every hierarchical level in organizations, may be why an employee leaves. Conflicts that cannot be resolved easily, in particular, trigger individual employee turnover. Staw (1980) states that this conflict-triggered but voluntary turnover may be seen as beneficial and not as a cost to the organization because it sometimes helps to "resolve deep-seated conflicts" between the conflicting parties and contributes to organizational morale.

Turnover and the resulting inflow of a new employee may be the primary source for heterogeneity in skills and mindsets within organizations (Staw, 1980). However, this consequence may be significantly more distinct with respect to collective turnover.

Taxonomy of Individual Turnover Consequences

Table 1 is a taxonomic overview of individual turnover consequences, derived from the discussion on the consequences of individual voluntary turnover. It identifies positive and negative consequences and distinguishes between those that occur on the *former employer side* and the *new employer side* because turnover refers both to entering and leaving the organization (Bluedorn, 1978). It distinguishes further between consequences for individuals (departing/arriving employees and their co-workers) and the organization as a whole.

	Consequences for		Positive	Negative
Former employer	Individual	Departing employee	Job attitude (Steers and Mowday, 1981)	Social capital loss (Pennings et al., 1998; Shaw et al., 2005) Job attitude (Steers and Mowday, 1981)
		Co-worker	Morale (Staw, 1980) Promotion opportunity (Staw, 1980)	Demoralization (Staw, 1980) Job Attitude (Steers and Mowday, 1981)
	Organization		Morale (Staw, 1980) Social capital gain (Corredoira and Rosenkopf, 2010)	Replacement costs (Mobley, 1982) Operational disruption (Staw, 1980) Human capital loss (Rosenkopf and Almeida, 2003) Knowledge spillover (Aime et al., 2010; Franco and Filson, 2006)

New employer	Individual	Arriving employee	Job satisfaction (Steers and Mowday, 1981) Mental health (Mobley, 1982) Organizational commitment (Steers and Mowday, 1981) Job performance (Muchinsky and Morrow, 1980; Steers and Mowday, 1981) Social capital gain (Dess and Shaw, 2001)	Anxiety (Feldman and Brett, 1983) Uncertainty (Ashford and Cummings, 1983)
		Co-worker	Organizational commitment (Yang, 2008)	Demoralization (Staw, 1980)
	Organization		Organizational performance (Grusky, 1960; Hausknecht and Trevor, 2011; Nyberg and Ployhart, 2013; Staw, 1980) Human capital gain (Carnahan et al., 2012; Franco and Filson, 2006) Social capital gain (Corredoira and Rosenkopf, 2010; Dess and Shaw, 2001) Skill and mindset heterogeneity (Staw, 1980) Knowledge spillover (Aime et al., 2010; Franco and Filson, 2006)	Training costs (Dess and Shaw, 2001) Recruitment costs (Dess and Shaw, 2001)

Table 1. Taxonomy of individual turnover consequences

Skills and Mindsets of IT personnel[1]

With respect to mindsets and skills, IT personnel differ among each other but also compared with the general population.

Lee et al. (1995) considers four categories of critical IT skills all IT employees should have: (1) technical specialties skills cover a range of IT technical specialties; (2) technology management skills

[1] There is inconsistency in terminology and conceptualization of the IT profession in the literature, but as the present study's focus does not include resolving this and providing a universal definition of IT personnel, the exact terms used in the studies referenced are used here.

answer the questions of where and how to deploy information technologies effectively and meet strategic business objectives profitably; (3) business functional skills cover both general business skills as well as knowledge of and ability to learn about business functions; and (4) interpersonal and management skills include the boundary-spanning role IT professionals must assume in organizations (Baroudi, 1985). Further, they suggest that industry needs a diverse cadre of IT professionals with knowledge and skills in technology, business operations, management, and interpersonal relationships to lead organizational integration and process reengineering activities effectively (Lee et al., 1995).

Research also found differences in the personality of employees across IT job types. System analysts and project managers are more conservative, logical, analytical, diligent, and ambitious than programmers, with stronger leadership tendencies, higher self-confidence, and greater self-esteem (Wynekoop and Walz, 1998). Consultants, programmers, system engineers, and system administrators differ in terms of personality and job-related attitudes (Eckhardt et al., 2016). Regarding the Big Five personality traits, system engineers are more open and conscientious, IT consultants more extroverted, programmers more neurotic, and system administrators more agreeable (Eckhardt et al., 2016).

Distinguishing between IS workers and IT professionals, Orlikowski and Baroudi (1988) reveal that IS workers have a distinct occupational culture and require a different knowledge and skills set than IT professionals. IT personnel also possess some unique characteristics that result from the rapid changes in technology that make existing skill sets obsolete (Agarwal and Ferratt, 2002). They have unique business competence in conceptual development and may influence IT-business partnerships. Knowledge sharing among IT professionals is socially driven (Tsai and Cheng, 2012); IT professionals play a potentially key role in transferring knowledge across organizational boundaries (Pawlowski and Robey, 2004). Today's IT professional should have project management knowledge, business domain knowledge, relationship skills, customer expertise, and problem-solving skills in addition to technical skills (Rutner et al., 2008). These skills are critical because they enable IT departments to work effectively with other departments, internal users, and external customers and suppliers (Gallagher et al., 2010). IT professionals possess a high need for learning and they have a strong desire to be challenged (Lee, 2000).

Prior studies have found no differences between IT and non-IT workers with respect to the motivators of productive work behavior and conclude that IT and non-IT employees at the same occupational level are not and should not be managed different (Ferratt and Short, 1986, 1988). Further, the boundaries dividing IT and non-IT personnel continue to become blurred and encompass non-IT personnel as they take on more primary tasks in the IS domain (Niederman et al., 2016).

However, other research has indicated some differences between IT workers and those in other domains (Bartol and Martin, 1982; Loh et al., 1995). IT technical/professionals and managers have lower social needs than non-IT individuals. Further, IT technical/professionals have a higher need for achievement than those in some other occupations (Bartol and Martin, 1982). Differences in personalities exist between IT employees and the general population, and between individuals in different job categories in the IS field (Wynekoop and Walz, 1998). IT personnel have established a distinct occupational culture within organizations, characterized by the use of technical jargon, the primary value of technical knowledge, feelings of superiority, and a general lack of formal rules (Guzman et al., 2008; Rao and Ramachandran, 2011). Further, IT professionals hold two kinds of specific human capital – firm-specific and IT-specific human capital (Josefek, Jr. and Kauffman, 2003) – that distinguish them from non-IT professionals (Mithas and Krishnan, 2008). IT-specific human capital is unique to IT jobs and less transferable to other occupations (Slaughter et al., 2007).

To conclude, IT personnel are different from the general workforce with respect to their mindsets and skills. Based on these observations from the literature, we propose that the voluntary turnover behavior of IT personnel may result in consequences different than those from non-IT turnover. We discuss this proposition in section 5, after presenting the current state of IT turnover consequences research in the following sections.

Scoping Literature Review

We conducted a multidisciplinary scoping review (Arksey and O'Malley, 2005; Paré et al., 2015), following the guidelines for literature reviews by Paré et al. (2016). Paré et al. (2016) guidelines emphasize the "transparency and systematicity" of literature reviews and provide a six-step framework for literature reviews. Scoping reviews focus on the breadth rather than the depth of coverage of the literature and provide an overview of the research gaps in the existing literature (Paré et al., 2015). As with systematic literature reviews (SLR), scoping reviews use rigorous methods that seek all literature relevant to the topic being studied. Methods need to be comprehensive and transparent, with the aim of being replicable. To make the review process as systematic and transparent as possible, we report all review steps in a detailed manner.

We chose a broad scope for this literature review and did not restrict our research to a single discipline. We included *Business Source Complete*, the American Economic Association's electronic bibliography *EconLit Database*, *SocINDEX,* and *PsycINFO* in our search via *EBSCOhost* to cover the most important management, economics, and applied psychology journals. Further, we included the digital libraries *ScienceDirect* (a reliable source for IS publications) and *JSTOR* (a reliable source for older publications). Leading IS journals (including the *Senior Scholars' Basket of Eight*[2]) and IS conference proceedings were covered by the *Association for Information Systems Research electronic library (AISeL)*, *EBSCOhost, ScienceDirect* and the *IEEE digital library* (for *Hawaii International Conference on System Sciences (HICSS) proceedings*). We included the proceedings of the *ACM Special Interest Group on Management of Information Systems Computers and Personnel Research* (*SIGMIS CPR*) in our search because this specific community investigates the needs, interests, and abilities of IT personnel. We limited the search population to peer-reviewed journal papers and conference proceedings, as they provide the necessary rigor and quality in scientific research. No limits were placed on the year of publication. We set September 1, 2016 as a publication cut-off date for articles to be included in our sample.

Table 2 shows the search pattern for potentially relevant articles. It is based on synonyms and differences in terminology and for two main keywords, turnover and employee, derived from brainstorming and from articles identified initially.

Title/Abstract/ Keywords (OR)		Title/Abstract/ Keywords (OR)
IS/IT, "IS/IT workforce," "IS/IT worker," "IS/IT professional," "IS/IT labor," "IS/IT personnel"	AND	Turnover, mobility, movement, flow, "job change", "career change"

Table 2. Search pattern

In the first step, we searched only for IT turnover studies in the aforementioned databases. The search results (without filtering) yielded 143 studies; after removing duplicates and studies that are irrelevant to IT turnover, 126 IT turnover studies remained. We further scanned the reference lists of these 126 studies and conducted reverse and forward searches to identify more relevant sources. In total, we found 153 IT

[2] See https://aisnet.org/?SeniorScholarBasket

turnover studies for detailed manual screening of the full text (see Figure 3 and Appendix A). Table 3 presents the criteria for further inclusion or exclusion of these 153 studies. Citavi[3] was used to file and manage retrieved papers and record decisions of the reviewer team.

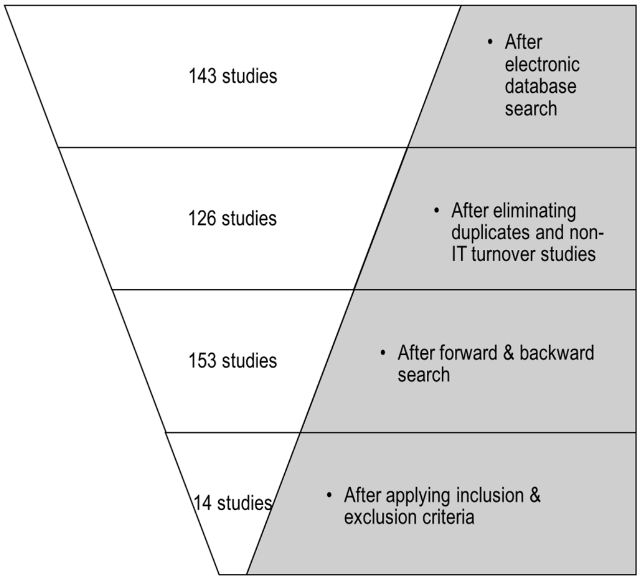


Figure 3. Stepwise search results of the keyword search

Inclusion Criteria	Exclusion Criteria
Published in English	Published in language other than English
Studies published up to September 1, 2016	IT turnover studies focused on cognitive precursors of turnover
Quantitative IT turnover studies focused on the consequences of individual voluntary turnover	IT turnover studies focused on the antecedents of turnover
Qualitative IT turnover studies focused on the consequences of individual voluntary turnover	IT turnover studies focused on collective turnover
Conceptual IT turnover studies focused on consequences of individual voluntary turnover	IT turnover studies about turnover into/out of the workforce or in entrepreneurship
Peer-reviewed journal papers and conference proceedings	IT turnover studies that mention consequences of individual voluntary turnover in an anecdotal way

Table 3. Inclusion and exclusion criteria for the literature review

[3] http://www.citavi.de/en/index.html

We also analyzed the studies to check how the publications are distributed in time (see Figure 4)[4]. We can see that IT turnover research started in the 1960s, and IT turnover consequences research in the 1980s. Just a few papers appeared from then until 1999. The last decade contains the majority of IT turnover and IT turnover consequences research. As already mentioned, the publication cut-off date was September 1st 2016, so it is possible that more papers were published in 2016 and were not indexed so far. In total, we found 14 of 153 IT turnover studies (see Appendix B) that deal about the consequences of individual voluntary IT turnover.

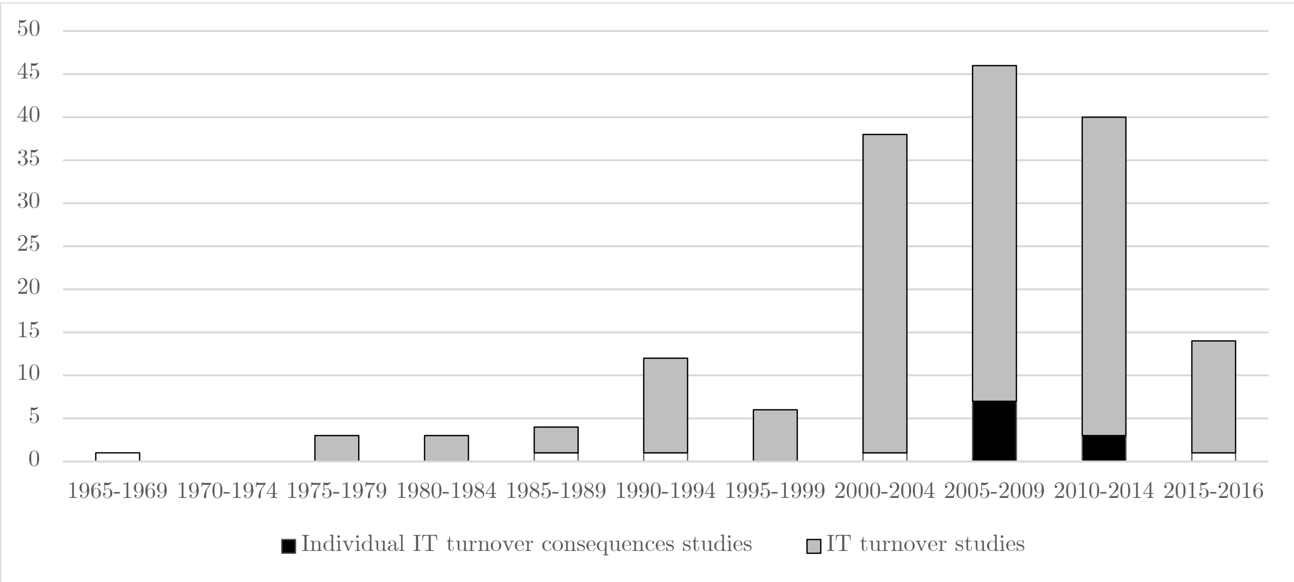


Figure 4. Distribution of studies found per year

We scanned and coded the 14 publications identified and then verified their inclusion in terms of quality. In literature reviews that describe or categorize the existing literature and identify gaps (e.g., scoping reviews), quality assessment in the strictest sense is neither essential nor a key requirement (Paré et al., 2016). However, we did perform a quality assessment to evaluate the rigor of the research presented in each publication. We assessed the 14 relevant studies using six quality criteria, as presented in Table 4. All questions had two possible answers: *Yes* or *No.* These answers were scored as follows: (1) or (0), respectively.

[4] A complete list of the reviewed IT turnover literature can be found in Appendix C.

Problem statement
Q1. Is the research objective sufficiently explained and well motivated?
Research design
Q2. Is the context of the study clearly stated? Q3. Is the research design prepared sufficiently?
Data (if applicable)
Q4. Is the data analysis used in the study adequately described?
Discussion & conclusion
Q5. Are the findings of the study clearly stated and supported by the results? Q6. Does the paper discuss limitations?

Table 4. Criteria for quality assessment

Even though the quality assessment criteria and their evaluation scales may be subjective, they do provide a common framework for comparing the selected papers. To ensure the validity of the 14 studies and the reliability of our findings, a study was included if its quality score exceeded 3 points (50% of the maximum quality score of an article: 6 points). Table 5 present the quality scores of the 14 selected articles. All 14 studies had acceptable quality scores and were included in this review.

Reference	Q1	Q2	Q3	Q4	Q5	Q6	Σ
Abdel-Hamid, 1989	1	1	1	1	1	0	**5**
Abdel-Hamid, 1992	1	1	1	1	1	1	**6**
Chowa, 2010	1	1	1	1	0	0	**4**
Dibbern et al., 2008	1	1	1	1	1	1	**6**
Freedman, 2008	1	1	1	1	1	1	**6**
Gopal et al., 2003	1	1	1	1	1	1	**6**
Gopal and Sivaramakrishnan, 2008	1	1	1	1	1	1	**6**
Hall et al., 2008	1	1	1	1	0	0	**4**
Izquierdo-Cortazar et al., 2009	1	1	1	1	1	1	**6**
Joseph et al., 2015	1	1	1	1	1	1	**6**
Mithas and Krishnan, 2008	1	1	1	1	1	1	**6**

Pee et al., 2008	1	1	1	1	1	1	**6**
Tambe and Hitt, 2014	1	1	1	1	1	1	**6**
Wu et al., 2014	1	1	1	1	1	1	**6**

Table 5. Quality assessment results

Research Results

This section presents and discusses the 14 studies focused on IT turnover behavior consequences that resulted from our literature review: seven that focus on the negative consequences of IT personnel turnover in IT projects; one that addresses the consequences of CIO turnover with respect to IT alignment; and three that address positive consequences for IT professionals in terms of salary growth. Two studies focus on productivity benefits from IT turnover.

Costs and knowledge loss in IT project management

Abdel-Hamid (1989) investigates how staff turnover, acquisition, and assimilation rates affect software development costs and schedules. The study's results indicate that these rates can increase a project's costs and duration up to 60 percent in worst-case scenarios, which suggests that turnover is critical for the successful development of software systems, as well as for the accurate estimation of software development costs and schedules. In a further study, Abdel-Hamid (1992) examines the impacts of managerial turnover/succession on software project performance. In particular, the study examines the staffing and cost/schedule tradeoff choices of successor project managers and compares them with the choices made by managers who run their projects from start to finish without interruption. The results indicate that managerial turnover can lead to a visible shift in cost/schedule tradeoff choices, affecting staff allocations and ultimately project performance in terms of both cost and duration. The resignation of heavily involved personnel may result in significant losses of vital undocumented information and necessitate the recollection of data related to requirements, for example. During the all project phases (programming, testing, and implementation), losses of key personnel can imply work overload to other project participants, and hence the project to fall behind schedule, lead to schedule and cost overruns.

Pee et al. (2008) study 151 development teams and claim that turnover can result in losing knowledge and weakens project performance when new members with different experiences join teams. The use of succession planning, knowledge repositories, and employee orientation programs to manage different types of organizational knowledge loss can effectively mitigate the detrimental effects of turnover.

Izquiredo-Cortazar et al. (2009) state that a turnover of senior developers leaves a knowledge gap that must be managed. Junior developers who replace a departed senior developer require some time to achieve the desired level of productivity. The time to learn how the project works results in significant productivity losses that are unavoidable when senior developers leave a project and are substituted for by others new to the project or to parts of the code they have to maintain.

The results of the Hall et al. (2008) study on the impact of staff turnover on software projects show that there may be a relationship between staff turnover and project success. However, the authors note that this relationship is not distinct and may vary across projects. Additionally the results suggest that project success is likely to be improved if staff turnover is controlled (Hall et al., 2008).

While specific project-related characteristics such as project team size and resource shortage largely explain contract choice in these projects, IT turnover seems to have no significant impact (Gopal et al., 2003). A few years later, Gopal and Sivaramakrishnan (2008) used the same data as Gopal et al. (2003) – 93 offshore projects completed by a leading Indian vendor – and could finally show that IT turnover from a project affects profits negatively. In the context of IS offshoring with the focus on software development and maintenance projects, IT turnover was found to increase client extra costs (Dibbern et al., 2008).

Productivity benefits

With respect to productivity, one can say that firms take advantage of IT investments of other firms from which they hire IT personnel. Firms derive significant productivity benefits from the hired IT personnel (Tambe and Hitt, 2014). Hiring IT personnel from a diverse set of firms can substantially improve firm productivity, likely due to the diverse and non-redundant information provided in a network with high diversity (Wu et al., 2014). With respect to productivity, one can say that firms take advantage of IT investments of other firms from which they hire IT personnel. Firms derive significant productivity benefits from the hired IT personnel (Tambe and Hitt, 2014). Hiring IT personnel from a diverse set of firms can substantially improve firm productivity, likely due to the diverse and non-redundant information provided in a network with high diversity (Wu et al., 2014).

IT alignment

The study by Chowa (2010) examines the organizational impact of CIO turnover and its effects on IT alignment. Results show that, on average, CIO turnover does not lead to fundamental change in the firm and has little impact on IT alignment. The authors argue that this may be due to the fact that CIO change occurs mainly as a result of low CIO performance. Further, CIO change affords the new CIO a chance to make changes that their predecessor was not able to make due to political and other relational issues (Chowa, 2010).

Salary growth and job status

Mithas and Krishnan (2008) study the influence of supply- and demand-side factors on the compensation of IT professionals and reveal that individual voluntary IT turnover leads to higher wages for the leaving employee. The former employer has to pay higher compensation for a new IT professional with the same skill and experience. The Mithas and Krishnan (2008) results confirm the widely prevalent notion that "job hopping is necessary for salary growth in the IT profession." Joseph et al.'s (2015) study reveals quite similar findings: IT professionals are able to increase job status and pay levels when moving to jobs within the IT profession. Similar results were found by Freedman (2008) in the context of the software publishing industry, where turnover behavior is an important aspect for salary growth.

Discussion and future research

In this study, we conducted a literature review of IT turnover literature focused on the consequences. Our goal was to determine what is known about the consequences of voluntary IT turnover behavior and identify research gaps. The results reveal several things.

First, there has not been a lot of research to date on the consequences of voluntary individual IT turnover behavior. While there is a growing body of literature on the turnover intentions of IT personnel and other direct and indirect antecedents of the actual turnover behavior, studies focused on the consequences are rare.

We identified 14 studies that discuss consequences of IT turnover. Most discuss potential negative consequences such as costs, low performance, and knowledge loss (see Table). Of these, the majority (seven) are in the context of IT project management; one study (Gopal et al., 2003) did not provide evidence for any consequences of IT turnover. Three studies mention the positive consequences (salary growth and job status) that turnover might imply. Further, only negative consequences for the former employer are identified (see Table 6).

<table>
<tr><th colspan="3">Consequences for</th><th>Positive</th><th>Negative</th></tr>
<tr><td rowspan="3">Former employer</td><td rowspan="2">Individual</td><td>Departing employee</td><td>GAP 1</td><td>GAP 5</td></tr>
<tr><td>Co-worker</td><td>GAP 2</td><td>Staff allocation (Abdel-Hamid, 1992)</td></tr>
<tr><td colspan="2">Organization</td><td>GAP 3</td><td>Project performance (Abdel-Hamid, 1992; Izquierdo-Cortazar et al., 2009; Pee et al., 2008)
Project costs (Abdel-Hamid, 1989, 1992; Dibbern et al., 2008; Gopal and Sivaramakrishnan, 2008; Hall et al., 2008; Pee et al., 2008)
Project schedule (Abdel-Hamid, 1989, 1992)
Knowledge loss (Izquierdo-Cortazar et al., 2009; Pee et al., 2008)
IT alignment (Chowa, 2010)
Compensation costs (Mithas and Krishnan, 2008)</td></tr>
<tr><td rowspan="3">New employer</td><td rowspan="2">Individual</td><td>Arriving employee</td><td>Salary growth (Freedman, 2008; Joseph et al., 2015; Mithas and Krishnan, 2008)
Job status (Joseph et al., 2015)</td><td>GAP 6</td></tr>
<tr><td>Co-worker</td><td>GAP 4</td><td>GAP 7</td></tr>
<tr><td colspan="2">Organization</td><td>Productivity benefits (Tambe and Hitt, 2014; Wu et al., 2014)</td><td>GAP 8</td></tr>
</table>

Table 6. Gap areas on research on consequences of individual voluntary IT turnover from our scoping review

Second, the review in section 2 shows that related disciplines consider consequences of individual turnover that have not been considered in the specific context of IT turnover. This is quite astonishing, because research on the consequences of employee turnover traces its beginning to work by Price (1977) many decades ago.

Third, we identified eight research gaps that should be addressed by IT turnover behavior research. *Gaps 1, 2,* and *3* (see Table 6) relate to the positive consequences for the departing employee, his or her former co-workers, and the former employer. The turnover process has positive consequences for the departing employee with respect to job attitudes (*Gap 1*). This also applies for IT professionals. Since turnover is the primary creator of promotion opportunities (Dalton and Todor, 1979) and IT personnel have a strong desire to be challenged (Lee, 2000), we assume that IT personnel may seize promotion opportunities to greater degrees than non-IT personnel (*Gap 2*). IT professionals cross departmental and organizational boundaries to complete projects and interact with many actors across these boundaries. Such interaction may play an important role in transferring knowledge across organizational boundaries (Baroudi, 1985; Pawlowski and Robey, 2004).

As studies from IS-related disciplines show, losing a highly skilled employee to a competitor may lead to access to external knowledge for the former employer (*Gap 3*) (Somaya et al., 2008). However, two studies of IT turnover consequences from our literature review show that knowledge loss is a negative consequence of individual voluntary IT turnover (Izquierdo-Cortazar et al., 2009; Pee et al., 2008). We assume that the boundary-spanning role of IT professionals may not only facilitate the flow of knowledge back to former employer firms (Aime et al., 2010; Madsen et al., 2003), but may also enhance the social networks of the former and new employer (*Gap 3*) (Corredoira and Rosenkopf, 2010). Hence, we need to study the relationship between former colleagues in IT-related occupations after an IT turnover. How does an individual IT turnover affect these relationships? Voluntary individual turnover disrupts established social relations with former colleagues (Pennings et al., 1998). When an IT professional leaves unexpectedly and remaining IT co-workers have to compensate by working longer hours, it may lead to bad moods and, as a consequence, remaining IT co-workers breaking off contact with the departed colleague (*Gap 5*). However, perhaps the former colleagues feel challenged after a colleague's turnover and the bad mood is less marked in an IT employee context than in a non-IT employee context.

Further, can we assume that IT employees have stronger ties to their former colleagues than other occupational groups? A firm might profit from these types of relationships with respect to social capital, but we cannot make a clear statement about this without an empirical investigation.

While several studies have investigated negative consequences of voluntary individual IT turnover, further research can investigate, for example, the consequences of IT turnover on training costs. Since most IT is standardized, firms are able to reduce training costs when recruiting IT professionals in comparison to non-IT personnel (Joseph et al., 2015). However, there is a higher turnover rate within the IT profession due to the fact that IT personnel have more job alternatives and seem to have a lower degree of organizational citizenship (Mithas and Krishnan, 2008). As a result, an employer would have to recruit IT personnel more often than non-IT personnel, leading to higher recruitment and training costs. Whether and to what extent recruitment and training costs of IT personnel are higher, lower, or the same should be clarified in further research.

Gap 4 addresses the positive consequences for the newcomer's co-workers. The arrival of a new IT employee might spur co-workers with respect to their productivity/performance and their organizational commitment.

Gaps 6, 7, and *8* concern the negative consequences for both individuals and the new employer. We reviewed the literature from disciplines such as applied psychology and organizational management to identify consequences that have not previously been researched in the context of IT personnel. One example is demoralization (Staw, 1980), a consequence that has not been researched in the IT turnover context. Anxiety and uncertainty on the individual level may not be that important in the IT context (*Gap 6*). The IS domain attracts employees that are highly motivated and flexible when it comes to learning new IT skills (e.g., learning a new programming language) (LeRouge et al., 2006). However, age and especially personality are important mediators that affect the consequences when a new IT employee starts his or her new job. The consequences of individual IT turnover on socialization should also be studied because they have significant influence on the socialization tactics utilized by employers (King and Xia, 2001) (*Gap 7*). We posit that IT professionals who voluntary change their employer are more willing to learn to deal with their new environments. This would match with their high need for learning and their strong desire to be challenged (Lee, 2000). However, this may not be valid for all IT-related occupations. "IT road warriors" (Ahuja et al., 2007) may experience socialization in a different way because they are socialized mostly by their clients and not by their employer organizations.

Training costs due to the heterogeneous skillset and the high need for growth and personal development and learning of IT professionals (Lee, 2000) are likely to be high compared to other occupations (*Gap 8*). However, we must consider the career stage of IT personnel. The turnover of early-career IT professionals leads to higher training costs for new employers and to lower replacement costs for former employer organizations. Lower training costs for new employers are implied for IT employees in mid- and late-career stages, but former employers incur higher replacement costs. However, turnover behavior creates opportunities for replacements, and such replacements may bring with them new knowledge, ideas, approaches, technologies, and working styles (Mobley, 1982).

We propose that researchers who examine the relevance of these consequences can reveal new insights into IT personnel, especially about their mindsets.

As with every study, there are some limitations that should be taken into account. While the evidence supports clear gaps and guidance for a new research agenda, our analysis and the taxonomy are limited to what is reported in the studies discussed. We restricted our review to articles we identified through our keyword search in a broad selection of electronic databases. Based on our scoping review of these articles, we are confident that we presented the research state of the art on the consequences of individual voluntary IT turnover. In addition, several consequences presented in this study are conceptual and provide opportunities for further empirical testing.

Further, one must consider that consequences have moderating variables. Staw (1980) presents an overview of potential moderating variables that affect the significance of several consequences. Moreover, it is not trivial to separate consequences for individuals from those for organizations.

Future IT turnover research concerning the consequences of individual voluntary turnover behavior should be aware of contexts other than IT project management, the different contexts of IT turnover, and the distinct forms of job mobility. Most studies on consequences of IT turnover are based in IT project management, ignoring these other important contexts. Further, we need to distinguish between employees who take similar jobs in the same occupation and those who take different jobs in different occupations (Joseph et al., 2015; Mowday et al., 1984). Joseph et al. (2015) propose a differentiation between turnover, turnaway-within, and turnaway-between forms of job mobility. Turnover may result in different consequences than turnaway-between job mobility. Hence, context is important. Studying turnover consequences implies that one has a clear understanding of the unit of analysis.

Beyond distinguishing the job mobility form when conducting research on IT turnover consequences, we also need to distinguish between different job profiles in the IT context. IT employees are not a homogeneous group (Lo and Riemenschneider, 2011), and different IT job types (programmer, IT consultant, system engineer, etc.) differ not only according to their job-related attitudes and turnover intentions but also according to their personalities (Eckhardt et al., 2016). Further, a researcher who builds a causal model of turnover and its consequences cannot take everything into account at the same time, and so exclusions of the individual or organizational effects are reasonable and make the task of model construction more manageable.

We believe this study contributes to research and practice in two ways. Only after weighing the pros and cons of turnover and its consequences can managers effectively manage their human resources. Turnover makes it crucially important that top management understand the effects, both direct and indirect, of turnover so they can plan it, or at least plan for it. Further, this study contributes to a better understanding of IT turnover in general and the consequences of IT turnover in particular.

In summary, more focused research is needed on the consequences of individual voluntary IT turnover. We have an incomplete understanding of the consequences of IT turnover for organizations and individuals. We hope our brief overview and research agenda provides a starting point for further IT turnover research not only regarding antecedents but the consequences as well.

Conclusion

The present study examined the current research state of individual voluntary IT turnover consequences. Building upon existing IT turnover literature from different disciplines, we revealed that there is a relatively vague understanding of the consequences of individual voluntary IT turnover. The findings, especially the research gaps we identified, have enabled us to establish a research agenda for IT turnover behavior research with a focus on the consequences. Thus, we contribute to the IT turnover research stream by emphasizing the relevance of studying the consequences of IT turnover behavior, not only its antecedents. In addition, our findings call attention to problems HR management must address and resolve. Only if HR management is aware of the antecedents and the consequences of IT turnover can managers effectively manage their human resources.

References

Abdel-Hamid, T. K. (1989). "A Study of Staff Turnover, Acquisition, and Assimilation and their Impact on Software Development Cost and Schedule," *Journal of Management Information Systems*. Vol. 6, No. 1: pp. 21–40.

Abdel-Hamid, T. K. (1992). "Investigating the Impacts of Managerial Turnover/Succession on Software Project Performance," *Journal of Management Information Systems*. Vol. 9, No. 2: pp. 127–144.

Agarwal, R., and Ferratt, T. W. (2002). "Enduring Practices for Managing IT Professionals," *Communications of the ACM*. Vol. 45, No. 9: pp. 73–79.

Ahuja, M. K., Chudoba, K. M., Kacmar, C. J., McKnight, D. H., and George, J. F. (2007). "IT Road Warriors: Balancing Work-family Conflict, Job Autonomy, and Work Overload to Mitigate Turnover Intentions," *MIS Quarterly*. Vol. 31, No. 1: pp. 1–17.

Aime, F., Johnson, S., Ridge, J. W., and Hill, A. D. (2010). "The Routine may be Stable but the Advantage is not: Competitive Implications of Key Employee Mobility," *Strategic Management Journal*. Vol. 31, No. 1: pp. 75–87.

Allen, T. D., Freeman, D. M., Russell, Joyce E. A., Reizenstein, R. C., and Rentz, J. O. (2001). "Survivor Reactions to Organizational Downsizing: Does Time ease the Pain?" *Journal of Occupational & Organizational Psychology*. Vol. 74, No. 2: pp. 145–164.

Arksey, H., and O'Malley, L. (2005). "Scoping Studies: Towards a Methodological Framework," *International Journal of Social Research Methodology*. Vol. 8, No. 1: pp. 19–32.

Ashford, S. J., and Cummings, L. L. (1983). "Feedback as an Individual Resource: Personal Strategies of Creating Information," *Organizational Behavior & Human Performance*. Vol. 32, pp. 370–398.

Baroudi, J. J. (1985). "The Impact of Role Variables on IS Personnel Work Attitudes and Intentions," *MIS Quarterly*. Vol. 9, No. 4.

Bartol, K. M., and Martin, D. C. (1982). "Managing Information Systems Personnel: A Review of the Literature and Managerial Implications," *MIS Quarterly*. Vol. 6, No. 5: pp. 49–70.

Bermiss, Y. S., and Murmann, J. P. (2015). "Who matters more? The Impact of Functional Background and Top Executive Mobility on Firm Survival," *Strategic Management Journal*. Vol. 36, No. 11: pp. 1697–1716.

Bluedorn, A. C. (1978). "A Taxonomy of Turnover," *Academy of Management Review*. Vol. 3, No. 3: pp. 647–651.

Bluedorn, A. C. (1982). "The Theories of Turnover: Causes, Effects, and Meaning," in *Perspectives in Organizational Sociology: Theory and Research*, S. Bacharach (ed.), Greenwich, CT, USA: JAI Press.

Borés, C., Saurina, C., and Torres, R. (2003). "Technological Convergence: A Strategic Perspective," *Technovation*. Vol. 23, No. 1: pp. 1–13.

Bureau of Labor Statistics, U.S. Department of Labor (2016). *Occupational Employment Statistics*, www.bls.gov/oes/. Accessed 1 September 2016.

Cable, D. M., Gino, F., and Staats, B. R. (2013). "Breaking Them in or Eliciting Their Best?: Reframing Socialization around Newcomers' Authentic Self-expression," *Administrative Science Quarterly*. Vol. 58, No. 1: pp. 1–36.

Carnahan, S., Agarwal, R., and Campbell, B. A. (2012). "Heterogeneity in Turnover: The Effect of Relative Compensation Dispersion of Firms on the Mobility and Entrepreneurship of Extreme Performers," *Strategic Management Journal*. Vol. 33, No. 12: pp. 1411–1430.

Chowa, C. K. (2010). "CIO Turnover, IS Alignment and Revolutionary Change", AMCIS 2010 Proceedings.

Corredoira, R. A., and Rosenkopf, L. (2010). "Should auld acquaintance be forgot? The Reverse Transfer of Knowledge through Mobility Ties," *Strategic Management Journal*. Vol. 31, No. 2: pp. 159–181.

Dalton, D. R., and Todor, W. D. (1979). "Turnover Turned Over: An Expanded and Positive Perspective," *Academy of Management Review*. Vol. 4, No. 2: pp. 225–235.

Dalton, D. R., Todor, W. D., and Krackhardt, D. M. (1982). "Turnover Overstated: The Functional Taxonomy," *Academy of Management Review*. Vol. 7, No. 1: pp. 117–123.

Dess, G. G., and Shaw, J. D. (2001). "Voluntary Turnover, Social Capital, and Organizational Performance," *Academy of Management Review*. Vol. 26, No. 3: pp. 446–456.

Dibbern, J., Winkler, J., and Heinzl, A. (2008). "Explaining Variations in Client Extra Costs Between Software Projects Offshored to India," *MIS Quarterly*. Vol. 32, No. 2: pp. 333–366.

Eckhardt, A., Laumer, S., Maier, C., and Weitzel, T. (2016). "The Effect of Personality on IT Personnel's Job-related Attitudes: Establishing a Dispositional Model of Turnover Intention across IT Job Types," *Journal of Information Technology*. Vol. 31, No. 1: pp. 48–66.

Feldman, D. C. (1976). "A Contingency Theory of Socialization," *Administrative Science Quarterly*. Vol. 21, No. 3: pp. 433–452.

Feldman, D. C., and Brett, J. M. (1983). "Coping with New Jobs: A Comparative Study of New Hires and Job Changers," *Academy of Management Journal*. Vol. 26, No. 2: pp. 258–272.

Ferratt, T. W., and Short, L. E. (1986). "Are Information Systems People Different? An Investigation of Motivational Differences," *MIS Quarterly*. Vol. 10, No. 4: pp. 377–387.

Ferratt, T. W., and Short, L. E. (1988). "Are Information Systems People Different? An Investigation of How They Are and Should be Managed," *MIS Quarterly*. Vol. 12, No. 3: pp. 427–443.

Festinger, L. (1962). *A Theory of Cognitive Dissonance*, Stanford: Stanford Univ. Press.

Franco, A. M., and Filson, D. (2006). "Spin-outs: Knowledge Diffusion through Employee Mobility," *RAND Journal of Economics*. Vol. 37, No. 4: pp. 841–860.

Freedman, M. L. (2008). "Job Hopping, Earnings Dynamics, and Industrial Agglomeration in the Software Publishing Industry," *Journal of Urban Economics*. Vol. 64, No. 3: pp. 590–600.

Gallagher, K. P., Kaiser, K. M., Simon, J. C., Beath, C. M., and Goles, T. (2010). "The Requisite Variety of Skills for IT professionals," *Communications of the ACM*. Vol. 53, No. 6: pp. 144–148.

García-Murillo, M., and MacInnes, I. (2003). "The Impact of Technological Convergence on the Regulation of ICT Industries," *Journal of Media Management*. Vol. 5, No. 5: pp. 57–67.

Gopal, A., and Sivaramakrishnan, K. (2008). "Research Note —On Vendor Preferences for Contract Types in Offshore Software Projects: The Case of Fixed Price vs. Time and Materials Contracts," *Information Systems Research*. Vol. 19, No. 2: pp. 202–220.

Gopal, A., Sivaramakrishnan, K., Krishnan, M. S., and Mukhopadhyay, T. (2003). "Contracts in Offshore Software Development: An Empirical Analysis," *Management Science*. Vol. 49, No. 12: pp. 1671–1683.

Grusky, O. (1960). "Administrative Succession in Formal Organizations," *Social Forces*. Vol. 39, pp. 105–115.

Guzman, I. R., Stam, K. R., and Stanton, J. M. (2008). "The Occupational Culture of IS/IT Personnel within Organizations," *DATA BASE for Advances in Information Systems*. Vol. 39, No. 1: pp. 33–50.

Hall, T., Beecham, S., Verner, J., and Wilson, D. (2008). "The Impact of Staff Turnover on Software Projects," SIGMIS CPR'08 Proceedings, Charlottesville, VA, USA, pp. 30–39.

Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., and Pierce, C. A. (2013). "Meta-analytic Review of Employee Turnover as a Predictor of Firm Performance," *Journal of Management*. Vol. 39, No. 3: pp. 573–603.

Hausknecht, J. P., and Trevor, C. O. (2011). "Collective Turnover at the Group, Unit, and Organizational Levels: Evidence, Issues, and Implications," *Journal of Management*. Vol. 37, No. 1: pp. 352–388.

Hom, P. W., and Griffeth, R. W. (1991). "Structural Equations Modeling Test of a Turnover Theory: Cross-sectional and Longitudinal Analyses," *Journal of Applied Psychology*. Vol. 76, No. 3: pp. 350–366.

Izquierdo-Cortazar, D., Robles, G., Ortega, F., and Gonzalez-Barahona, J. M. (2009). "Using Software Archaeology to Measure Knowledge Loss in Software Projects due to Developer Turnover," HICSS '09 Proceedings.

Josefek, R. A., Jr., and Kauffman, R. J. (2003). IT Human Capital and the Information Systems Professional's Decision to Leave the Company.

Joseph, D., Ang, S., and Slaughter, S. A. (2015). "Turnover or Turnaway? Competing Risks Analysis of Male and

Female IT Professionals' Job Mobility and Relative Pay Gap," *Information Systems Research*. Vol. 26, No. 1: pp. 145-164.

Joseph, D., Fong Boh, W., Ang, S., and Slaughter, S. A. (2012). "The Career Paths Less (or More) Traveled: A Sequence Analysis of IT Career Histories, Mobility Patterns, and Career Success," *MIS Quarterly*. Vol. 36, No. 2: pp. 427-452.

Joseph, D., Kok-Yee Ng, Koh, C., and Soon Ang (2007). "Turnover of Information Technology Professionals: A Narrative Review, Meta-Analytic Structural Equation Modeling, and Model Development," *MIS Quarterly*. Vol. 31, No. 3: pp. 547–577.

King, R. C., and Xia, W. (2001). "Retaining IS Talent in the New Millennium: Effects of Socialization on IS Professionals' Role Adjustment and Organizational Attachment," SIGCPR '01 Proceedings, San Diego, CA, USA, pp. 144-156.

Krackhardt, D., and Porter, L. W. (1986). "The Snowball Effect: Turnover Embedded in Communication Networks," *Journal of Applied Psychology*. Vol. 71, No. 1: pp. 50–55.

Krackhardt, D. M., and Porter, L. W. (1985). "When Friends Leave: A Structural Analysis of the Relationship between Turnover and Stayers' Attitudes," *Administrative Science Quarterly*. Vol. 30, No. 2: pp. 242–261.

Lacity, M. C., Iyer, V. V., and Rudramuniyaiah, P. S. (2008). "Turnover Intentions of Indian IS Professionals," *Information Systems Frontiers*. Vol. 10, No. 2: pp. 225–241.

Lee, D. M., Trauth, E. M., and Farwell, D. (1995). "Critical Skills and Knowledge Requirements of IS Professionals: A Joint Academic/Industry Investigation," *MIS Quarterly*. Vol. 19, No. 3: pp. 313–340.

Lee, P. C. (2000). "Turnover of Information Technology Professionals: A Contextual Model," *Accounting, Management and Information Technologies*. Vol. 10, No. 2: pp. 101–124.

Lee, T. W., and Mitchell, T. R. (1994). "An Alternative Approach: The Unfolding Model of Employee Turnover," *Academy of Management Review*. Vol. 19, No. 1: pp. 51–89.

LeRouge, C., Nelson, A., and Blanton, J. Ellis (2006). "The Impact of Role Stress Fit and Self-esteem on the Job Attitudes of IT Professionals," *Information & Management*. Vol. 43, No. 8: pp. 928–938.

Lindbeck, A., and Snower, D. J. (2001). "Insiders versus Outsiders," *Journal of Economic Perspectives*. Vol. 15, No. 1: pp. 165–188.

Lo, J. (2015). "The Information Technology Workforce: A Review and Assessment of Voluntary Turnover Research," *Information Systems Frontiers*. Vol. 17, No. 2: pp. 387–411.

Lo, J., and Riemenschneider, C. (2011). "Heterogeneity of IT Employees: An Analysis of a Model of Perceived Organizational Support by Job Type," *DATA BASE for Advances in Information Systems*. Vol. 42, No. 3: pp. 71–95.

Loh, L., Sankar, C. S., and Yeong, W. Y. (1995). "Job Orientation, Perceptions, and Satisfaction a Study of Information Technology Professionals in Singapore," *Information & Management*. Vol. 29, No. 5: pp. 239–250.

Louis, M. R. (1980). "What Newcomers Experience in Entering Unfamiliar Organizational Settings," *Administrative Science Quarterly*. Vol. 25, pp. 226–251.

Madsen, T. L., Mosakowski, E., and Zaheer, S. (2003). "Knowledge Retention and Personnel Mobility: The Nondisruptive Effects of Inflows of Experience," *Organization Science*. Vol. 14, No. 2: pp. 173–191.

March, J. G., and Simon, H. A. (1958). *Organizations*, New York, NY, USA: Wiley.

Mithas, S., and Krishnan, M. S. (2008). "Human Capital and Institutional Effects in the Compensation of Information Technology Professionals in the United States," *Management Science*. Vol. 54, No. 3: pp. 415–428.

Mobley, W. H. (1977). "Intermediate Linkages in the Relationship Between Job Satisfaction and Employee Turnover," *Journal of Applied Psychology*. Vol. 62, No. 2: pp. 237–240.

Mobley, W. H. (1982). Employee Turnover: Causes, Consequences, and Control, Addison Wesley.

Mobley, W. H., Griffeth, R. W., Hand, H. H., and Meglino, B. M. (1979). "Review and Conceptual Analysis of the Employee Turnover Process," *Psychological Bulletin*. Vol. 86, No. 3: pp. 493–522.

Mourmant, G., Gallivan, M. J., and Kalika, M. (2009). "Another Road to IT Turnover: The Entrepreneurial

Path," *European Journal of Information Systems*. Vol. 18, No. 5: pp. 498–521.

Mourmant, G., and Voutsina, K. (2010). "From IT Employee to IT Entrepreneur: The Concept of IT Entrepreneurial Epiphany,"

Mowday, R. T., Koberg, C. S., and McArthur, A. W. (1984). "The Psychology of the Withdrawal Process: A Cross-Validational Test of Mobley's Intermediate Linkages Model of Turnover in Two Samples," *Academy of Management Journal*. Vol. 27, No. 1: pp. 79–94.

Mowday, R. T., Porter, L. W., and Steers, R. M. (1982). *Employee-Organization Linkages: The Psychology of Commitment, Absenteeism, and Turnover*, New York: Academic Press.

Muchinsky, P. M., and Morrow, P. C. (1980). "A Multidisciplinary Model of Voluntary Employee Turnover," *Career Adaptability*. Vol. 17, No. 3: pp. 263–290.

Muller, C. W., and Price, J. L. (1989). "Some Consequences of Turnover: A Work Unit Analysis," *Human Relations*. Vol. 42, No. 5: pp. 389–403.

Niederman, F., Ferratt, T. W., and Trauth, E. M. (2016). "On the Co-Evolution of Information Technology and Information Systems Personnel," *DATA BASE for Advances in Information Systems*. Vol. 47, No. 1: pp. 29–50.

Niederman, F., Sumner, M., and Maertz Jr. , Carl P. (2007). "Testing and Extending the Unfolding Model of Voluntary Turnover to IT Professionals," *Human Resource Management*. Vol. 46, No. 3: pp. 331–347.

Nyberg, A. J., and Ployhart, R. E. (2013). "Context-emergent Turnover (CET) Theory: A Theory of Collective Turnover," *Academy of Management Review*. Vol. 38, No. 1: pp. 109–131.

Orlikowski, W. J., and Baroudi, J. J. (1988). "The Information Systems Profession: Myth or Reality?" *Office Technology and People*. Vol. 4, No. 1: pp. 13–30.

Paré, G., Tate, M., Johnstone, D., and Kitsiou, S. (2016). "Contextualizing the Twin Concepts of Systematicity and Transparency in Information Systems Literature Reviews," *European Journal of Information Systems*.

Paré, G., Trudel, M.-C., Jaana, M., and Kitsiou, S. (2015). "Synthesizing Information Systems Knowledge: A Typology of Literature Reviews," *Information & Management*. Vol. 52, No. 2: pp. 183–199.

Pawlowski, S. D., and Robey, D. (2004). "Bridging User Organizations: Knowledge Brokering and the Work of Information Technology Professionals," *MIS Quarterly*. Vol. 28, No. 4: pp. 645–672.

Pee, L. G., Tham, Z.-C., Kankanhalli, A., and Tan, G. W. (2008). "Turnover in Information Systems Development Projects - Managing Forgetting,"

Pennings, J. M., Lee, K., and van Witteloostuijn, A. (1998). "Human Capital, Social Capital, and Firm Dissolution," *Academy of Management Journal*. Vol. 41, No. 4: pp. 425–440.

Porter, L. W., Crampon, W. J., and Smith, F. J. (1976). "Organizational Commitment and Managerial Turnover: A Longitudinal Study," *Organizational Behavior & Human Performance*. Vol. 15, No. 1: pp. 87–98.

Price, J. L. (1977). *The Study of Turnover*, Ames, Iowa: Iowa State Univ. Press.

Rao, V. S., and Ramachandran, S. (2011). "Occupational Cultures of Information Systems Personnel and Managerial Personnel: Potential Conflicts," *Communications of the Association for Information Systems*. Vol. 29, pp. 582–604.

Rosenkopf, L., and Almeida, P. (2003). "Overcoming Local Search Through Alliances and Mobility," *Management Science*. Vol. 49, No. 6: pp. 751–766.

Rutner, P. S., Hardgrave, B. C., and McKnight, D. H. (2008). "Emotional Dissonance and the Information Technology Professional," *MIS Quarterly*. Vol. 32, No. 3: pp. 635–652.

Seddon, P. B., and Scheepers, R. (2012). "Towards the Improved Treatment of Generalization of Knowledge Claims in IS Research: Drawing General Conclusions from Samples," *European Journal of Information Systems*. Vol. 21, No. 1: pp. 6–21.

Shapiro, D. L., Hom, P., Shen, W., and Agarwal, R. (2016). "How Do Leader Departures Affect Subordinates Organizational Attachment?: A 360-Degree Relational Perspective," *Academy of Management Review*. Vol. 41, No. 3: pp. 479–502.

Shaw, J. D., Duffy, M. K., Johnson, J. L., and Lockhart, D. E. (2005). "Turnover, Social Capital Losses, and

Performance," *Academy of Management Journal*. Vol. 48, No. 4: pp. 594–606.

Slaughter, S. A., Ang, S., and Fong Boh, W. (2007). "Firm-specific Human Capital and Compensation Organizational Tenure Profiles: An Archival Analysis of Salary Data for IT," *Human Resource Management*. Vol. 46, No. 3: pp. 373–394.

Somaya, D., Williamson, I. O., and Lorinkova, N. (2008). "Gone but not Lost: The Different Performance Impacts of Employee Mobility between Cooperators versus Competitors," *Academy of Management Journal*. Vol. 51, No. 5: pp. 936–953.

Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Fischbach, K., Schlagwein, D., and Schoder, D. (2014). "Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors," *Academy of Management Proceedings*.

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., and Schoder, D. (2016). "Business Model Development, Founders' Social Capital and the Success of Early Stage Internet Start-ups: A Mixed-method Study," *Information Systems Journal*. Vol. 26, No. 5: pp. 421–449.

Staw, B. M. (1980). "The Consequences of Turnover," *Journal of Occupational Behavior*. Vol. 1, No. 4: pp. 253–273.

Steers, R. M. (1977). "Antecedents and Outcomes of Organizational Commitment," *Administrative Science Quarterly*. Vol. 22, No. 1: pp. 46–56.

Steers, R. M., and Mowday, R. T. (1981). "Employee Turnover and Post-decision Accommodation Processes," in *Research in Organizational Behavior*, L. L. Cummings and B. M. Staw (eds.), Greenwich, CT, USA: JAI Press.

Tambe, P., and Hitt, L. M. (2014). "Job Hopping, Information Technology Spillovers, and Productivity Growth," *Management Science*. Vol. 60, No. 2: pp. 338–355.

Ton, Z., and Huckman, R. S. (2008). "Managing the Impact of Employee Turnover on Performance: The Role of Process Conformance," *Organization Science*. Vol. 19, No. 1: pp. 56–68.

Tsai, M.-T., and Cheng, N.-C. (2012). "Understanding Knowledge Sharing between IT Professionals – An Integration of Social Cognitive and Social Exchange Theory," *Behaviour & Information Technology*. Vol. 31, No. 11: pp. 1069–1080.

Wu, L., Jin, F., and Hitt, L. M. (2014). "Are All Spillovers Created Equal? A Network Perspective on IT Labor Movements," ICIS 2014 Proceedings.

Wynekoop, J. L., and Walz, D. B. (1998). "Revisiting the Perennial Question: Are IS People Different?" *DATA BASE for Advances in Information Systems*. Vol. 29, No. 2: pp. 62–72.

Yang, J.-T. (2008). "Effect of Newcomer Socialisation on Organisational Commitment, Job Satisfaction, and Turnover Intention in the Hotel Industry," *Service Industries Journal*. Vol. 28, No. 4: pp. 429–443.

Appendix A: Results of the Keyword Search Grouped by Database

Database	Keyword search	- Duplicates - Non IT turnover studies	Applying inclusion & exclusion criteria
EBSCOhost	32	31	3
ScienceDirect	10	9	0
AISeL	29	24	3
ACM DL (DATA BASE & SIGMIS CPR Proceedings)	54	47	1
IEEE DL (HICSS Proceedings)	8	8	1
JSTOR	10	7	2
Forward and backward search	-	27	4
∑	143	153	14

Appendix B: List of IT Turnover Consequences Studies

Author(s)	Title	Year	Outlet
Abdel-Hamid, T. K.	A Study of Staff Turnover, Acquisition, and Assimilation and their Impact on Software Development Cost and Schedule	1989	Journal of Management Information Systems
Abdel-Hamid, T. K.	Investigating the impacts of managerial turnover/succession on software project performance	1992	Journal of Management Information Systems
Chowa, C. K.	CIO turnover, IS alignment and revolutionary change	2010	AMCIS Proceedings
Dibbern, J.; Winkler, J.; Heinzl, A.	Explaining Variations in Client Extra Costs Between Software Projects Offshored to India	2008	MIS Quarterly
Freedman, M. L.	Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry	2008	Journal of Urban Economics
Gopal, A.; Sivaramakrishnan, K.; Krishnan, M. S.; Mukhopadhyay, T.	Contracts in Offshore Software Development: An Empirical Analysis	2003	Management Science

Gopal, A.; Sivaramakrishnan, K.	Research Note - On Vendor Preferences for Contract Types in Offshore Software Projects: The Case of Fixed Price vs. Time and Materials Contracts	2008	Information Systems Research
Hall, T.; Beecham, S.; Verner, J.; Wilson, D.	The impact of staff turnover on software projects: the importance of understanding what makes software practitioners tick	2008	SIGMIS CPR Proceedings
Izquierdo-Cortazar, D.; Robles, G.; Ortega, F.; Gonzalez-Barahona, J. M.	Using software archaeology to measure knowledge loss in software projects due to developer turnover	2009	HICSS Proceedings
Joseph, D.; Ang, S.; Slaughter, S. A.	Turnover or turnaway? Competing risks analysis of male and female IT professionals' job mobility and relative pay gap	2015	Information Systems Research
Mithas, S.; Krishnan, M. S.	Human capital and institutional effects in the compensation of information technology professionals in the United States	2008	Management Science
Pee, L. G.; Tham, Z.; Kankanhalli, A.; Tan, G. W.	Turnover in information systems development projects - Managing forgetting	2008	PACIS Proceedings
Tambe, P.; Hitt, L. M.	Job hopping, information technology spillovers, and productivity growth	2014	Management Science
Wu, L.; Jin, F.; Hitt, L. M.	Are all spillovers created equal? A network perspective on IT labor movements	2014	ICIS Proceedings

Appendix C: List of IT Turnover Studies

ID	Author(s)	Title	Year	Outlet
1	Abdel-Hamid, Tarek K.	A Study of Staff Turnover, Acquisition, and Assimilation and their Impact on Software Development Cost and Schedule	1989	Journal of Management Information Systems
2	Abdel-Hamid, Tarek K.	Investigating the impacts of managerial turnover/succession on software project performance	1992	Journal of Management Information Systems
3	Adya, Monica P.	Work alienation among IT workers: a cross-cultural gender comparison	2008	SIGMIS CPR '08 Proceedings 2008
4	Agarwal, Ritu; Ferratt, Thomas W.; De, P.	An experimental investigation of turnover intentions among new entrants in IT	2007	ACM SIGMIS Data Base
5	Ahuja, Manju K.; Chudoba, K. M.; George, J. F.; Kacmar, C.; McKnight, H.	Overworked and isolated? Predicting the effect of work-family conflict, autonomy, and workload on organizational commitment and turnover of virtual workers	2002	Proceedings of the 35th Annual Hawaii International Conference on System Science
6	Ahuja, Manju K.; Chudoba, K. M.; Kacmar, Charles J.; McKnight, D. Harrison; George, Joey F.	IT road warriors: Balancing work - family conflict, job autonomy, and work overload to mitigate turnover intentions	2007	MIS Quarterly
7	Allen, Myria W.; Armstrong, Deborah J.; Reid, Margaret F.; Riemenschneider, Cynthia K.	IT employee retention: employee expectations and workplace environments	2009	SIGMIS CPR '08 Proceedings 2009
8	Ang, Soon; Slaughter, Sandra A.	Turnover of information technology professionals: The effects of internal labor market strategies.	2004	ACM SIGMIS Data Base
9	Armstrong, Deborah J.; Brooks, Nita G.; Riemenschneider, Cynthia K.	Exhaustion from information system career experience: Implications for turn-away intention	2015	MIS Quarterly
10	Armstrong, Deborah J.; Riemenschneider, Cynthia K.; Allen, Myria W.; Reid, Margaret F.	Advancement, voluntary turnover and women in IT: A cognitive study of work–family conflict	2007	Information & Management
11	Armstrong, Deborah J.; Riemenschneider, Cynthia K.; Reid, Margaret F.; Nelms, Jason E.	Challenges and barriers facing women in the IS workforce: How far have we come?	2011	SIGMIS CPR '11 Proceedings 2011
12	Bartol, Kathryn M.	Turnover among DP personnel: A casual analysis	1983	Communications of the ACM
13	Bartol, Kathryn M.; Martin, David C.	Managing the consequences of DP turnover: A human resources planning perspective	1983	SIGCPR '83 The Proceedings 1983
14	Bidwell, Matthew J.; Briscoe, Forrest	The dynamics of interorganizational careers	2010	Organization Science
15	Black, Sandra E.; Lynch, Lisa M.	How to Compete	2001	Review of Economics and Statistics
16	Bound, John; Demirci, Murat; Khanna, Gaurav; Turner, Sarah	Finishing degrees and finding jobs: US higher education and the flow of foreign IT workers	2015	Innovation Policy and the Economy

17	Bradford, P. A.; Cottrell, L. R.	Factors influencing business data processors turnover - A comparative case history	1977	ACM '77 Proceedings 1977
18	Burley, Diana L.; Guzman, Indira R.; Pandit, Gayatri	Will they stay? Turnover intentions of future federal cyber corps members	2010	SIGMIS CPR '10 Proceedings 2010
19	Carayon, P.; Schoepke, J.; Hoonakker, Peter; Haims, M.; Brunette, M.	Evaluating causes and consequences of turnover intention among IT workers: the development of a questionnaire survey	2006	Behaviour & Information Technology
20	Carr, Christopher; Jones, Brian	Organizational culture and the antecedents of turnover in high-stress IT jobs	2001	AMCIS 2001 Proceedings
21	Chang, Christina Ling-Hsing	The study of the turnover of MIS professionals—The gap between Taiwanese and US societies	2010	International Journal of Information Management
22	Chang, Christina Ling-Hsing; Jiang, James J.; Klein, Gary; Chen, Houn-Gee	Career anchors and disturbances in job turnover decisions – A case study of IT professionals in Taiwan	2012	Information & Management
23	Chang, Christina Ling-Hsing; Lin, I-Chun	Career Anchors, National Culture and Leave Intent of MIS Professionals in Taiwan	2008	PACIS 2008 Proceedings
24	Chang, Sug-In	Work role stressors and turnover intentions: A study of IT personnel in South Korea	2008	Zeitschrift für Personalforschung / German Journal of Research in Human Resource Management
25	Chowa, Charles K.	CIO turnover, IS alignment and revolutionary change	2010	AMCIS 2010 Proceedings
26	Chung, Q.; Wagner, William; Luo, Wenhong	Loyalty of IT workforce in the digial economy	2001	AMCIS 2001 Proceedings
27	D'Mello, Marisa; Sahay, Sundeep	"I am kind of a nomad where I have to go places and places"... Understanding mobility, place and identity in global software work from India	2007	Information and Organization
28	Darais, Karin M.; Nelson, Kay; Rice, Sarah; Buche, Mari	Identifying the enablers and barriers of information technology personnel transition	2001	ICIS 2001 Proceedings
29	Dibbern, Jens; Winkler, Jessica; Heinzl, Armin	Explaining Variations in Client Extra Costs Between Software Projects Offshored to India	2008	MIS Quarterly
30	Dickson, C. M.	In-house recruiting - one answer to programmer force losses	1969	SIGCPR '69 Proceedings 1969
31	Dinger, Michael; Grover, Varun; Thatcher, Jason B.	The Impact of Embeddedness on IT Worker Behavior	2009	AMCIS 2009 Doctoral Consortium
32	Dinger, Michael; Thatcher, Jason B.; Stepina, Lee P.; Craig, Kevin	The grass is always greener on the other side: A test of present and alternative job utility on IT professionals' turnover	2012	IEEE Transactions on Engineering Management
33	Dinger, Michael; Thatcher, Jason B.; Treadway, Darren; Stepina, Lee; Breland, Jacob	Does Professionalism Matter in the IT Workforce? An Empirical Examination of IT Professionals	2015	Journal of the Association for Information Systems
34	Dittrich, J. E., Couger, J. D.; Zawacki, R. A.	Perceptions of equity, job satisfaction, and intention to quit among data processing personnel	1985	Information & Management

35	Eckhardt, Andreas; Laumer, Sven; Maier, Christian; Weitzel, Tim	The effect of personality on IT personnel's job-related attitudes: Establishing a dispositional model of turnover intention across IT job types	2016	Journal of Information Technology (Palgrave Macmillan)
36	Egan, T. M.; Yang, B.; Bartlett, K. R.	The effects of organizational learning culture and job satisfaction on motivation to transfer learning and turnover intention	2004	Human Resource Development Quarterly
37	Ertürk, Alper	Influences of HR Practices, Social Exchange, and Trust on Turnover Intentions of Public IT Professionals	2014	Public Personnel Management
38	Ferratt, Thomas W.; Agarwal, Ritu; Brown, Carol V.; Moore, Jo Ellen	IT human resource management configurations and IT turnover: Theoretical synthesis and empirical analysis	2005	Information Systems Research
39	Ford, Valerie F.; Swayze, Susan; Burley, Diana L.	An Exploratory Investigation of the Relationship between Disengagement, Exhaustion and Turnover Intention among IT Professionals Employed at a University	2013	Information Resources Management Journal
40	Freedman, Matthew L.	Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry	2008	Journal of Urban Economics
41	Gallivan, Mike; McLean, Ephraim; Moore, Jo Ellen; Roldan, Malu	Panel: Dimensions of mobility in the IT profession examining "Turnover Culture" and "Staying Behavior"	2001	SIGCPR '01 Proceedings 2001
42	Garden, A.-M.	Behavioural and organisational factors involved in the turnover of high tech professionals	1988	ACM SIGCPR Computer Personnel
43	Ge, Chunmian; Zou, Xiao	Human resource flow within software industry: A firm-level investigation	2013	PACIS 2013 Proceedings
44	Ghapanchi, Amir Hossein; Aurum, Aybuke	Antecedents to IT personnel's intentions to leave: A systematic literature review	2011	Journal of Systems and Software
45	Gill, Brian; Pidduck, Anne Banks	IT staffing and retention: a success story	2001	SIGCPR '01 Proceedings 2001
46	Gopal, Anandasivam; Sivaramakrishnan, Konduru	Research Note —On Vendor Preferences for Contract Types in Offshore Software Projects	2008	Information Systems Research
47	Gopal, Anandasivam; Sivaramakrishnan, Konduru; Krishnan, M. S.; Mukhopadhyay, Tridas	Contracts in Offshore Software Development: An Empirical Analysis	2003	Management Science
48	Guha, Sumana; Chakrabarti, Subhendu	Employee turnover: A study on information technology sector	2014	Journal of Business & Management
49	Guha, Sumana; Chakrabarti, Subhendu	Differentials in Attitude and Employee Turnover Propensity: A Study of Information Technology Professionals	2016	Global Business & Management Research
50	Guimaraes, T.; Igbaria, Magid	Determinants of turnover intentions: Comparing IC and IS personnel	1992	Information Systems Research
51	Hagel, William J. von; Miller, Leslie A.	Precipitating events leading to voluntary employee turnover among information technology professionals	2011	Journal of Leadership Studies
52	Hall, Tracy; Beecham, Sarah; Verner, June; Wilson, David	The impact of staff turnover on software projects: the importance of understanding what makes software practitioners tick	2008	SIGMIS CPR '08 Proceedings 2008
53	Hester, A. J.; Moore, Jo Ellen; Yager, S. E.	The role of voice in retention of IT workers: Paving the higher road	2014	Proceedings of the 47th Hawaii International

				Conference on System Science
54	Hsu, M. K.; Jiang, James J.; Klein, Gary; Tang, Z.	Perceived career incentives and intent to leave	2003	Information & Management
55	Hunter, M. G.; Tan, F. B.; Tan, B. C. Y.	Voluntary turnover of information systems professionals: A cross-cultural investigation	2008	Journal of Global Information Management
56	Igbaria, Magid; Greenhaus, Jeffrey H.	Determinants of MIS employees' turnover intentions: a structural equation model	1992	Communications of the ACM
57	Igbaria, Magid; Guimaraes, T.	Exploring differences in employee turnover intentions and its determinants among telecommuters and non-telecommuters	1999	Journal of Management Information Systems
58	Igbaria, Magid; Meredith, G.; Smith, D. C.	Predictors of intention of IS professionals to stay with the organization in South Africa	1994	Information & Management
59	Igbaria, Magid; Siegel, S. R.	The reasons for turnover of information systems personnel	1992	Information & Management
60	Izquierdo-Cortazar, D.; Robles, G.; Ortega, F.; Gonzalez-Barahona, J. M.	Using software archaeology to measure knowledge loss in software projects due to developer turnover	2009	Proceedings of the 42nd Hawaii International Conference on System Sciences
61	Jiang, James J.; Klein, Gary	A discrepancy model of information system personnel turnover	2002	Journal of Management Information Systems
62	Josefek, R.A., Jr.; Kauffman, R. J.	Five degrees of separation: A human capital model of employment-related decisionmaking in the information technology workforce	1999	Proceedings of the 32nd Hawaii International Conference on System Sciences
63	Josefek, Robert A.; Kauffman, Robert J.	Nearing the threshold: An economics approach to pressure on information systems professionals to separate from their employer	2003	Journal of Management Information Systems
64	Joseph, Damien; Ang, Soon	The threat-rigidity model of professional obsolescence and its impact on occupational mobility behaviors of IT professionals	2001	ICIS 2001 Proceedings
65	Joseph, Damien; Ang, Soon	Turnover of IT professionals: A quantitative analysis of the literature	2003	SIGCPR '03 Proceedings 2003
66	Joseph, Damien; Ang, Soon; Slaughter, Sandra A.	Turnover or turnaway? Competing risks analysis of male and female IT professionals' job mobility and relative pay gap	2015	Information Systems Research
67	Joseph, Damien; Ang, Soon; Slaughter, Sandra A.	Examining the role of general and firm-specific human capital in predicting IT professionals' turnover behaviors	2006	SIGCPR '06 Proceedings 2006
68	Joseph, Damien; Boh, Wai Fong; Ang, Soon; Slaughter, Sandra A.	The career paths less (or more) traveled: A sequence analysis of IT career histories, mobility patterns, and career success	2012	MIS Quarterly
69	Joseph, Damien; Ng, Kok-Yee; Koh, Christine; Ang, Soon	Turnover of information technology professionals: A narrative review, meta-analytic structural equation modeling, and model development	2007	MIS Quarterly
70	Ketler, Karen; Smith, Robert D.	Placement, performance and turnover of information systems professionals: Implications for HRIS	1993	SIGCPR '93 Proceedings 1993
71	Kim, Soonhee	Factors affecting state government information technology employee turnover intentions	2005	American Review of Public Administration

72	Kim, Soonhee	The Impact of Human Resource Management on State Government IT Employee Turnover Intentions	2012	Public Personnel Management
73	King, Ruth C.; Xia, Weidong	Retaining IS Talents in the new millennium: Effects of socialization on IS professionals' role adjustment and organizational attachment	2001	SIGCPR '01 Proceedings 2001
74	Klenke, Karin; Kievit, Karen-Ann	Predictors of leadership style, organizational commitment and turnover of information systems professionals	1992	SIGCPR '92 Proceedings 1992
75	Korsakienė, Renata; Stankevičienė, Asta; Šimelytė, Agnė; Talačkienė, Milda	Factors driving turnover and retention of information technology professionals	2015	Journal of Business Economics & Management
76	Krishnan, Sandeep K.; Singh, Manjari	Outcomes of intention to quit of Indian IT professionals	2010	Human Resource Management
77	Kym, Hyogun; Park, Won-Woo	The effect of cultural fit/misfit on the productivity and turnover of IS personnel	1992	SIGCPR '92 Proceedings 1992
78	Lacity, M. C.; Iyer, V. V.; Rudramuniyaiah, P. S.	Turnover intentions of Indian IS professionals	2008	Information Systems Frontiers
79	Laumer, Sven; Maier, Christian; Weitzel, Tim; Eckhardt, Andreas	The implementation of large-scale information systems in small and medium-sized enterprises - A case study of work-and health-related consequences	2012	Proceedings of the 45th Hawaii International Conference on System Science
80	Lee, Kyootai; Joshi, Kailash; Bae, Mueun	An Investigation of the Influence of the IS Context on the Determinants of Turnover Intentions in Korea	2010	Journal of Organizational Computing and Electronic Commerce
81	Lee, Patrick	The impact of role variables on turnover intentions of information technology professionals: An examination of moderating effects	2001	AMCIS 2001 Proceedings
82	Lee, Patrick C. B.	Career plateau and professional plateau: Impact on work outcomes of information technology professionals	1999	ACM SIGCPR Computer Personnel
83	Lee, Patrick C. B.	The social context of turnover among information technology professionals	2002	SIGCPR '02 Proceedings 2002
84	Lee, Patrick; Ang, Soon; Slaughter, Sandra A.	Turning Over Versus Turning Away of Information Systems Professionals	1997	ICIS 1997 Proceedings
85	Lerouge, Cynthia; Nelson, Anthony; Blanton, J. Ellis	The impact of role stress fit and self-esteem on the job attitudes of IT professionals	2006	Information & Management
86	Lin, Tung-Ching; Chang, Christina Ling-Hsing; Tsai, Wen-Chin	The influences of knowledge loss and knowledge retention mechanisms on the absorptive capacity and performance of a MIS department	2016	Management Decision
87	Lo, Janice; Riemenschneider, Cynthia K.	Heterogeneity of IT Employees: An Analysis of a Model of Perceived Organizational Support by Job Type	2011	ACM SIGMIS Data Base
88	Longenecker, Clinton O.; Scazzero, Joseph A.	The turnover and retention of IT managers in rapidly changing organizations	2003	Information Systems Management
89	MacCrory, Frank; Choudhary, Vidyanand; Pinsonneault, Alain	Research Note-Designing Promotion Ladders to Mitigate Turnover of IT Professionals	2016	Information Systems Research

90	Maier, Christian; Laumer, Sven; Eckhardt, Andreas; Weitzel, Tim	Analyzing the impact of HRIS implementations on HR personnel's job satisfaction and turnover intention	2013	Journal of Strategic Information Systems
91	Maier, Christian; Laumer, Sven; Eckhardt, Andreas; Weitzel, Tim	Who really quits?: A longitudinal analysis of voluntary turnover among IT personnel	2015	ACM SIGMIS Data Base
92	McKnight, D. Harrison; Phillips, Brandis; Hardgrave, Bill C.	Which reduces IT turnover intention the most: Workplace characteristics or job characteristics?	2009	Information & Management
93	Meland, Havard; Waage, Rolf Petter; Sein, Maung K.	The other side of turnover: Managing IT personnel strategically	2005	SIGCPR '05 Proceedings 2005
94	Mithas, Sunil; Krishnan, M. S.	Human capital and institutional effects in the compensation of information technology professionals in the United States	2008	Management Science
95	Moore, Jo Ellen	One road to turnover: An examination of work exhaustion in technology professionals	2000	MIS Quarterly
96	Moore, Jo Ellen	Illuminating the other road: The role of voice in IT turnover	2011	Proceedings of the 44th Hawaii International Conference on System Science
97	Moore, Jo Ellen; Burke, L. A.	How to turn around 'Turnover Culture' in IT	2002	Communications of the ACM
98	Mourmant, Gaëtan	Adapting and extending the unfolding model of voluntary job turnover to IS entrepreneurs'	2008	SIGMIS CPR '08 Proceedings 2008
99	Mourmant, Gaëtan	A necessary clarification of the unfolding model of voluntary turnover	2009	SIGMIS CPR '08 Proceedings 2009
100	Mourmant, Gaëtan; Gallivan, Mike	How personality type influences decision paths in the unfolding model of voluntary job turnover: an application to IS professionals	2007	SIGMIS CPR '07 Proceedings 2007
101	Mourmant, Gaëtan; Gallivan, Mike; Kalika, Michel	Another road to IT turnover: the entrepreneurial path	2009	European Journal of Information Systems
102	Mourmant, Gaëtan; Voutsina, Katerina	From IT employee to IT entrepreneur: The concept of IT entrepreneurial epiphany	2010	ICIS 2010 Proceedings
103	Nelson, Kay; Darais, Karin M.; Buche, Mari; Rice, Sarah	Identifying the constructs of IT personnel transition	2001	AMCIS 2001 Proceedings
104	Niederman, Fred A.; Sumner, Mary R.	Job turnover among MIS professionals: an exploratory study of employee turnover	2001	SIGCPR '01 Proceedings 2001
105	Niederman, Fred A.; Sumner, Mary R.	Decision paths affecting turnover among information technology professionals	2003	SIGCPR '03 Proceedings 2003
106	Niederman, Fred A.; Sumner, Mary R.; Carl P. Maertz, [JR.]	An analysis and synthesis of research related to turnover among IT personnel	2006	SIGCPR '06 Proceedings 2006
107	Niederman, Fred A.; Sumner, Mary R.; Maertz Jr., Carl P.	Testing and extending the unfolding model of voluntary turnover to IT professionals	2007	Human Resource Management
108	Palmer, Jonathan; Speier, Cheri; Buckley, Michael; Moore, Jo Ellen	Recruiting and retaining IS personnel: Factors influencing employee turnover	1998	SIGCPR '98 Proceedings 1998

109	Paré, Guy; Lalonde, Patrick	The impact of human resources practices on IT personnel commitment, citizenship behaviors, and turnover intentions	2000	ICIS 2000 Proceedings
110	Paré, Guy; Tremblay, Michel	The influence of high-involvement human resources practices, procedural justice, organizational commitment, and citizenship behaviors on information technology professionals' turnover intentions	2007	Group & Organization Management
111	Pee, L. G.; Kankanhalli, A.; Tan, G. W.; Tham, G. Z.	Mitigating the Impact of Member Turnover in Information Systems Development Projects	2014	IEEE Transactions on Engineering Management
112	Pee, Loo Geok; Tham, Zhi-Choong; Kankanhalli, Atreyi; Tan, Gek Woo	Turnover in information systems development projects - Managing forgetting	2008	PACIS 2008 Proceedings
113	Pelley, Lee; Kappelman, Leon; Vanacek, Michael	The impact of computer aided systems engineering on employee attitudes, job commitment and turnover	1992	SIGCPR '92 Proceedings 1992
114	Quan, Jim; Cha, H.	IT certifications, outsourcing and information systems personnel turnover	2010	Information Technology & People
115	Ramos, Eduardo; Joia, Luiz Antonio	An investigation into turn-away among information technology professionals in Brazil	2013	The Journal of High Technology Management Research
116	Reich, Blaize Horner; Kaarst-Brown, Michelle Lynn	"Seeding the Line": Understanding the transition from IT to Non-IT careers	1999	MIS Quarterly
117	Riemenschneider, Cynthia K.; Allen, Myria W.; Reid, Margaret F.	Potential antecedents to the voluntary turnover intentions of women working in information technology	2002	AMCIS 2002 Proceedings
118	Riemenschneider, Cynthia K.; Armstrong, Deborah J.; Allen, Myria W.; Reid, Margaret F.	What I'm not willing to share: A discussion of turnover and barriers to promotion with women IT workers	2004	AMCIS 2004 Proceedings
119	Riemenschneider, Cynthia K.; Armstrong, Deborah J.; Allen, Myria W.; Reid, Margaret F.	Barriers facing women in the IT work force	2006	ACM SIGMIS Data Base
120	Rigas, P. P.	A model of turnover intention among technically-oriented information systems professionals	2009	Information Resources Management Journal
121	Rouse, P. D.	Voluntary turnover related to information technology professionals: A review of rational and instinctualmodels	2001	International Journal of Organizational Analsyis
122	Rutner, Paige S.; Hardgrave, Bill C.; McKnight, D. Harrison	Emotional dissonance and the information technology professional	2008	MIS Quarterly
123	Ryan, Sherry; Prybutok, Victor; Zhang, Xiaoni	Job satisfaction and turnover among IT professionals: A cognitive dissonance approach	2006	AMCIS 2006 Proceedings
124	SamGnanakkan, S.	Mediating role of organizational commitment on HR practices and turnover intention among ICT professionals	2010	Journal of Management Research
125	SharifHeravi, Mona Ghafourian; Shahidi, Seyed Emad;	Investigating the relationships between leadership style and personnel turnover intention in IT companies in Iran	2010	SIGMIS CPR '10 Proceedings 2010

	Mahmood, Nik Hasnaa Nik			
126	Shih, Sheng-Pao; Jiang, James J.; Klein, Gary; Wang, Eric	Learning demand and job autonomy of IT personnel: Impact on turnover intention	2011	Computers in Human Behavior
127	Slaughter, Sandra A.; Ang, Soon	Internal labor market strategies and turnover of information technology professionals	2002	SIGCPR '02 Proceedings 2002
128	Smith, D. C.; Speight, H. L.	Antecedents of turnover intention and actual turnover among information systems personnel in South Africa	2006	SIGCPR '06 Proceedings 2006
129	Sumner, Mary R.; Niederman, Fred A.	The impact of gender differences on job satisfaction, job turnover, and career experiences of information systems professionals	2002	SIGCPR '02 Proceedings 2002
130	Sumner, Mary R.; Yager, Susan	Career orientation of IT personnel	2004	SIGCPR '04 Proceedings 2004
131	Tambe, Prasanna	Inter-industry IT spillovers after the Dot-Com bust	2012	ICIS 2012 Proceedings
132	Tambe, Prasanna; Hitt, Lorin M.	Job hopping, information technology spillovers, and productivity growth	2014	Management Science
133	Tambe, Prasanna; Hitt, Lorin M.	Now IT's personal: Offshoring and the shifting skill composition of the U.S. information technology workforce	2012	Management Science
134	Tan, Margaret; Igbaria, Magid	Exploring the status of the turnover and salary of information technology professionals in Singapore	1993	SIGCPR '93 Proceedings 1993
135	Tan, Margaret; Igbaria, Magid	Turnover and remuneration of information technology professionals in Singapore	1994	Information & Management
136	Tanniru, Mohan R.; Taylor, Susan M.	Causes of turnover among data processing professionals - Some preliminary findings	1981	SIGCPR '81 Proceedings 1981
137	Thatcher, Jason B.; Liu, Y.; Stepina, Lee P.; Goodman, J. M.; Treadway, Darren C.	IT worker turnover: An empirical examination of intrinsic motivation	2006	ACM SIGMIS Data Base
138	Thatcher, Jason B.; Liu, Yongmei; Stepina, Lee P.	The role of the work itself: an empirical examination of intrinsic motivation's influence on IT workers attitudes and intentions	2002	SIGCPR '02 Proceedings 2002
139	Thatcher, Jason B.; Stepina, Lee P.	Information technology worker turnover: An integrative model and empirical test	2001	ICIS 2001 Proceedings
140	Thatcher, Jason B.; Stepina, Lee P.; Boyle, Randall J.	Turnover of information technology workers: Examining empirically the influence of attitudes, job characteristics, and external markets	2002	Journal of Management Information Systems
141	Uruthirapathy, Aareni A.; Grant, Gerald G.	The influence of job characteristics on IT and non-IT job professional's turnover intentions	2015	Journal of Management Development
142	Uzoka, Faith-Michael E.; Mgaya, Klodwig V.; Shemi, Alice P.; Kitindi, Ernest G.; Akinnuwesi, Boluwaji A.	Stay or quit: IT personnel turnover in botswana	2011	SIGMIS CPR '11 Proceedings 2011
143	Voutsina, Katerina; Mourmant, Gaëtan	A literature review of the IT entrepreneurial turnover conditions: A comparison between U.S., France and Greece	2010	SIGMIS CPR '10 Proceedings 2010
144	Wang, Xinwei; Teo, Hock-Hai; Yang, Xue	Turnover intentions of IT employees in non-IT organizations: Effects of organizational and professional identification	2010	SIGMIS CPR '10 Proceedings 2010
145	Willoughby, Theodore C.	Computing personnel turnover: A review of the literature	1977	ACM '77 Proceedings 1977

146	Windeler, J.; Moore, Jo Ellen; Riemenschneider, Cynthia K.	Viewing turnover through a wide-angle lens: Conceptualizing locality turnover	2015	Proceedings of the 48th Hawaii International Conference on System Science
147	Wingreen, Stephen C.; Blanton, J. Ellis; Kittner, Marcy L.	The relationship between IT professionals' individual factors, training climate fit, and turnover intentions	2002	SIGCPR '02 Proceedings 2002
148	Wolfe, Jack M.	Personnel turnover rates	1977	ACM '77 Proceedings 1977
149	Wu, Lynn; Jin, Fujie; Hitt, Lorin M.	Are all spillovers created equal? A network perspective on IT labor movements	2014	ICIS 2014 Proceedings
150	Wynne, Lee A.; Ferratt, Thomas W.; Biros, David P.	Career anchors of United States Air Force information systems workers: A turnover predictor?	2002	SIGCPR '02 Proceedings 2002
151	Yetton, Philip; martin, Andrew; Sharma, Rajeev; Johnston, Kim	A model of information systems development project performance	2000	Information Systems Journal
152	Yu, Yiqing; Benlian, A.; Hess, T.	An empirical study of volunteer members' perceived turnover in open source software projects	2015	Proceedings of the 45th Hawaii International Conference on System Science
153	Zhang, Xiaoni; Ryan, Sherry D.; Prybutok, Victor R.; Kappelman, Leon	Perceived obsolescence, organizational embeddedness, and turnover of it workers	2012	ACM SIGMIS Data Base

Paper III

Bibliographic data	Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K. & Schoder, D. (2015). *Business Model Development, Founders' Social Capital, and the Success of Early-Stage Internet Startups: A Mixed-Method Study.* In: Information Systems Journal (ISJ), Vol. 26, No. 5, pp.421-449, https://doi.org/10.1111/isj.12073
Status of publication	Published as completed research paper.
Publication type	Journal article
Version & Copyright notice	Accepted version This is the peer reviewed version of the following article: *Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K. & Schoder, D. (2015). Business Model Development, Founders' Social Capital, and the Success of Early-Stage Internet Startups: A Mixed-Method Study. In: Information Systems Journal (ISJ), Vol. 26, No. 5, pp.421-449*, which has been published in final form at *https://doi.org/10.1111/isj.12073.* This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.
Previous versions	Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., Schoder, D. (2013). *The Impact of Structural Embeddedness on Funding Success of Early Stage Web Startups.* In: DIBME 2013 Pre-ECIS Workshop on the Digitization in Business Models and Entrepreneurship, Utrecht, The Netherlands. Zylka, M. P., Spiegel, O., Fischbach, K. (2013). *Employee Turnover in the IT Industry and its Impact on Social Capital and Success of Startup Companies.* In: XXXIII Sunbelt Social Networks Conference of the International Network for Social Network Analysis (INSNA), Hamburg, Germany

Business Model Development, Founders' Social Capital, and the Success of Early-Stage Internet Startups: A Mixed-Method Study

Information technology (IT) and entrepreneurship are more closely related than ever. The Internet in particular inspires the current 'generation startup'. While some early-stage Internet startups have quickly become major successes, others fail to secure required follow-up funding and collapse. In this paper, we build on and extend the emerging business model research stream with the aim of better understanding the differences between successful and unsuccessful early-stage Internet startups. In the qualitative first part of our mixed-method study, 17 expert informant interviews reveal that Internet startup business models are in permanent flux, continually changed and adapted by founders, who identify their professional social network (i.e., their social capital) as a critically important factor for developing the business model and ultimately making their startups successful. In the quantitative second part of the study, we test this claim based on a social network analysis (SNA) of 70 Internet startups and their 145 founders. We find strong support for the critical importance of the founders' social capital for early-stage Internet startup success. The findings of this study advance our understanding of the relationship between founders' social capital, the development of business models, and the success of early-stage Internet startups.

Keywords: Internet startups, social capital, business models, entrepreneurs, mixed-method research.

Introduction

Information technology (IT) and entrepreneurship are having a love affair, because 'IT is the magic ingredient that inspires and most often enables contemporary entrepreneurial endeavours' (Del Giudice & Straub, 2011, p. III). This link between IT and entrepreneurship is especially salient in the context of Internet startups (McKnight *et al.*, 2002; Kollmann, 2006; Serarois-Tarrés *et al.*, 2006), which leverage Internet technologies and the economics of digital products (e.g., network effects) to gain a competitive advantage through early market entry (Grover & Saeed, 2004). Some Internet startups have quickly become major successes that influence the lives of millions: Google, Skype, Facebook – and more recently Dropbox, WhatsApp, Instagram, and many others. Many more Internet startups, however, have failed and gone out of business.

The success of Internet startups depends heavily on the founders' ability to develop their business models – which may take years of continuous adaptation (Poole, 2001; Andries & Debackere, 2007). Successful Internet startups are able to adapt their business models continually, while maintaining liquidity (through funding) to support their operations (Grover & Saeed, 2004). Securing 'second round' or 'Series A' funding (i.e., a major investment by venture capitalists to support growth) is widely considered as a factor that distinguishes successful from unsuccessful startups (Burton *et al.*, 2002; Davila *et al.*, 2003; Baum & Silverman, 2004). According to one estimate, only 40 per cent of early-stage Internet startups are able to secure required Series A funding (Taylor, 2012).

The aim of this paper is to understand better why some early-stage Internet startups achieve this success while others do not (with success being considered as achieving Series A funding). We build on and extend two distinct research streams – business model research and entrepreneurship research – to gain new insights into the 'inner workings' of early-stage Internet startups.

Of late, the business model concept has seen growing attention from research and practice (for recent reviews see Lambert & Davidson, 2013; Veit *et al.*, 2014). A business model is 'the design of organizational structures to enact a commercial opportunity' (George & Bock, 2011, p. 99) – in short, it describes how organizations deliver value to their customers. The business model concept is helpful for gaining a more detailed understanding of the 'inner workings' and nature of a business (Al-Debei & Avison, 2010). However, substantive empirical IS research on business models is lacking (Zott *et al.*, 2011; Veit *et al.*, 2014). Further, the link between the business model and firm performance, as well as the process of business model change and development, are poorly understood (Al-Debei & Avison, 2010; George & Bock, 2011). What has been reported in prior research is that startups often change and adapt their business models (Shirky, 2008; McGrath, 2010; Teece, 2010). This may be especially relevant for early-stage Internet startups, where technological change and limited experience with reference products and services may require substantial changes to the business model. Internet startup success and the ability to perform continual adaptation of the business model are closely linked (Grover & Saeed, 2004). This leads us to our first research question: *What are the characteristics of early-stage Internet startup business models and how are these actually developed?* We answer this question primarily through the qualitative first part of our study.

Some literature claims that business model development depends heavily on the founders' abilities (Andries & Debackere, 2007; Trimi & Berbegal-Mirabent, 2012). Entrepreneurial research often focuses on the founders (i.e., entrepreneurs) rather than on the business model. The central question here is: why is it that some people are able to discover and exploit business opportunities while others are not (Shane & Venkataraman, 2000)? Studies of individual entrepreneurs in other empirical contexts have found that the founders' social capital (i.e., their social networks and personal relationships) is helpful in identifying new business opportunities (Ardichvili *et al.*, 2003; Bhagavatula *et al.*, 2010). Social capital may be particularly relevant for early-stage Internet startups because it can provide founders with access to critical knowledge, resources, and investors. This is a perception commonly held in the 'Internet startup scene', but it is based on anecdotal evidence. In light of the above claims in the literature and the common perception in practice (evident through our qualitative investigation as well as practitioners' writings), we ask a second research question: *Are early-stage Internet startups with better-connected founders more successful?* We answer this question through both parts of our study, adding to the qualitative a specific empirical test in the quantitative second part of our study.

To answer these two research questions, we designed a mixed-method study (Teddlie & Tashakkori, 2009; Creswell & Plano Clark, 2011). We conducted expert informant interviews with 17 Internet startup founders in the first part of our study and learned that founders consider the business model to be in permanent flux in the early startup stage. Further, they attributed key importance to their social capital for developing the business model and making their startups successful. Building on this finding, in the second part of our study we analysed a unique dataset covering 70 Internet startups and their 145 founders using social network analysis (SNA) techniques and found that better-connected founders were much more likely to create successful startups. Based on the overall findings of the study, we conclude that business models of early-stage Internet startups are highly dynamic and that founders leverage their social capital in developing the business model. Well-connected founders are most critical for the success of early-stage Internet startups.

The remainder of the paper is structured as follows. First, we review related empirical and theoretical work. We then explain our research design. After that, we present the findings and analysis, first of our qualitative study and then of our quantitative study. We discuss the theoretical and practical implications of the findings before concluding our study.

Related Works and Theoretical Foundations

Because of the interdisciplinary nature of the research, our review of the related literature needs to cover several different (and largely disconnected) literature streams. We therefore focus on a selection of key papers that inform our study and provide references to dedicated reviews of the different streams for the interested reader.

The business model concept (it is not a 'theory' as such) is critically important for this study and offers a theoretical approach for studying the 'black box' of what happens within firms. This is in contrast to traditional industry analysis that treats firms within one industry as essentially equal. The business model concept builds on two theoretical lenses. The market-based view draws from Porter's value chain concept (1985) and the extended notions of strategic positioning (Morris *et al.*, 2006). Further, the business model concept builds on the resource-based view of the firm (Wernerfelt, 1984; Barney, 1986) and related theories such as the knowledge-based view (Nonaka, 1994) and the dynamic capabilities/absorptive capacities perspective (Cohen & Levinthal, 1990; Teece *et al.*, 1997; Zahra & George, 2002). While it is beyond the scope of this paper to discuss the claims of these theories in detail, they share a common theme in that they aim to explain differences in firm performance by achieving sustainable competitive advantages through a firm's strategic market position, the various resources available to the firm, and the firm's ability to use these resources effectively and efficiently. The business model perspective acknowledges the role of resources and capabilities (Hedman & Kalling, 2003) but goes beyond this 'possession' perspective to consider explicitly other factors relevant to firm performance such as external value networks and the design (structure) of value creation (e.g., Al-Debei & Avison, 2010; George & Bock, 2011; Zott *et al.*, 2011). Hence, we find the business model perspective an appropriate, contemporary framework to study the differences between Internet startups.

The emergence of the business model concept is closely related to the emergence of the Internet in the mid-1990s (Zott *et al.*, 2011). Since then, it has seen increasing and substantial attention, especially in the fields of entrepreneurship (George & Bock, 2011), strategic management (Zott *et al.*, 2011), and information systems research (Al-Debei & Avison, 2010). Only recently, several efforts have been made to bring a unifying structure (Krumeich *et al.*, 2012) and a frame of reference to the different business model research streams (Al-Debei & Avison, 2010; George & Bock, 2011).

As noted above, we follow the straightforward business model definition of George & Bock (2011), who emphasize the role of opportunity discovery, ideation, and enactment. Ultimately, business models create value by exploiting the underlying opportunity (George & Bock, 2011). The value proposition describes the benefits customers can expect from products and services (Osterwalder *et al.*, 2014), and it is widely considered as a key business model element (see Afuah & Tucci, 2001; Chesbrough & Rosenbloom, 2002; Morris *et al.*, 2005; Al-Debei & Avison, 2010; Burkhart *et al.*, 2011; Krumeich *et al.*, 2012, among others). Finding the right fit between what a company offers and what customers want (i.e., the product-market fit) is crucial to any business model (Osterwalder *et al.*, 2014).

Al-Debei & Avison (2010) identify three additional common elements of business models in the literature: value architecture (i.e., the configuration of assets, resources, and core competencies), value network (i.e., the relationships to customers and other stakeholders such as partners, suppliers, etc.), and value finance (i.e., the financial setup in terms of costing, pricing, and revenue structure). However, despite these recent efforts, the exact conceptualization of the business model remains a matter of ongoing discussions in academia (e.g., Burkhart *et al.*, 2011; Lambert & Davidson, 2013). Further large-scale empirical analyses and conceptual work are needed to advance our understanding of business models (Al-Debei & Avison, 2010; Burkhart *et al.*, 2011; Veit *et al.*, 2014).

How is the business model relevant for explaining the performance and success of the firm? Previous studies have analysed the impact of different business model configurations on financial performance, (e.g., revenue growth, profitability, market capitalisation), and equity growth, as well as non-financial performance, such as resilience in challenging markets and the ability to provide social value to stakeholders (Lambert & Davidson, 2013). In the entrepreneurial context, prior work has primarily looked at more mature firms in their later stage of development: It has been shown that firms with business models that include novel elements (e.g., innovative combinations of products, services, and information) outperform those with business models that do not include such novel elements (Zott & Amit, 2007). From a dynamic perspective, initial empirical evidence suggests that business model adaptation and firm performance are positively related in new businesses (Andries & Debackere, 2007). The latter result corresponds to the argument that business models are inherently dynamic (Hedman & Kalling, 2003; MacInnes, 2005; Osterwalder *et al.*, 2005; Al-Debei & Avison, 2010), especially so in the IT industry (Trimi & Berbegal-Mirabent, 2012). It is considered a best practice to challenge and revise the business model continually (Blank, 2005; Osterwalder *et al.*, 2010; Ries, 2011). That is, rather than having a static business model, startups need to be dynamic, and sometimes even consider pivots – 'structured course corrections' (Ries, 2011, p. 103).

The goal of business model development is to find viable value propositions and create organizational structures that allow for exploiting the underlying opportunities (Blank, 2005; George & Bock, 2011; Osterwalder *et al.*, 2014), which in the end determines success or failure of a (startup) firm. In some cases this business model development can take years of continuous adaptation (Poole, 2001; Andries & Debackere, 2007). Startups go through several stages of development, which usually include modifications of their business models. Kollmann (2006) describes three distinct development phases: 'early stage', 'expansion stage', and 'later stage'. As noted before, we focus on the early stage of startups (also called the startup stage, or seed stage), which can be described as 'the state of a company when it has just been incorporated and its founders are developing their product or service' (NVCA, 2013, p. 74). The early stage begins with the initial work on the startup (i.e., the founders begin to work on an initial idea) and typically ends with either the startup receiving Series A funding or the startup being discontinued. During the early stage, companies are typically funded by founders' savings, friends and family, angel investors, or seed funding.

Measuring the success or performance of entrepreneurial firms is not trivial. For instance, Andries & Debackere (2007) measure their long-term survival rate, while Zott & Amit (2007) refer to stock market values. However, these are measures for companies in their later stages. Early-stage startups are different: unlike publicly traded companies, they are not required to publish company data. Revenues (which often do not exist) or growth rates are widely considered not fully representative of an early-stage startup's real value.

What defines success for an early-stage startup? Typically, founders, investors, and market observers consider it a success if the startup is evaluated positively by (i.e., receives funding from) a venture capitalist (VC) (Baum & Silverman, 2004), which 'confirms the quality of the company and decreases the uncertainty about its potential success. ... The credibility associated with a funding event – emanating from the information available to the VC firm as well as its reputation – gives a strong signal about the quality of the startup' (Davila *et al.*, 2003, p. 692). That is why, in this study, we follow prior research in that we consider the success of early-stage startups as obtaining Series A funding through a VC (Burton *et al.*, 2002; Davila *et al.*, 2003; Baum & Silverman, 2004).

Having argued for the importance of business models and their dynamism, as well as for Series A funding as an indicator of success for early-stage startups, we now must turn to the core question of what makes an early-stage startup successful.

At its core, entrepreneurship is about people and their processes around discovery, evaluation, and exploration of business opportunities (Shane & Venkataraman, 2000). Technology is critical, but it is ancillary to the people who take advantage of the business opportunities around them (Blank, 2005). From this perspective, business models do not work on their own; rather, they need to be implemented and continually adapted by capable entrepreneurs (Chesbrough, 2010). In fact, VCs regard the management team as most important when they evaluate investment opportunities (Muzyka *et al.*, 1996). In particular, entrepreneurs need to be able to explore and exploit technological capabilities and business opportunities (Blank, 2005; Blank & Dorf, 2012). The search for a viable value proposition is a continual, iterative process of designing and testing prototypes (Osterwalder *et al.*, 2014). A central finding of entrepreneurial research is that the abilities of people in otherwise comparable situations vary widely in these regards (Shane & Venkataraman, 2000). These performance differences of entrepreneurs (or, in the case of startups, founders) may also strongly affect the success of their startups.

Typically, entrepreneurial research has studied the question of success and performance differences by examining the personal abilities of individuals (Zhao *et al.*, 2010) and human capital (for a review see Unger *et al.*, 2011). Less understood is the role of the founders' social capital, which Stam *et al.* (2014) found significant in explaining the performance of small firms. Granovetter (1985) was among the first to argue that all economic action should be considered as embedded in social structures. Research on social capital focuses on explaining the performance and success of any individual actor embedded in (and as a function of) the surrounding social structures (Borgatti & Foster, 2003).

Social capital has been defined as 'the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit' (Nahapiet & Ghoshal, 1998, p. 243). Social capital may have significantly greater impacts than human capital for small businesses (for a review, see Stam *et al.*, 2014). Researchers have singled out repeatedly the benefit of social capital for firms that results from the ability of individuals with high social capital to identify new business opportunities for the firm (Ardichvili *et al.*, 2003; Bhagavatula *et al.*, 2010). Based on that argument, we infer that social capital is also relevant for the success of early-stage Internet startups, and likely even more so than human capital. This underscores the relevance of this study as it integrates the business model perspective (firm level) with research on the human and social capital (individual level) of entrepreneurs/founders in the context of early-stage consumer Internet startups.

Research on business models has certainly acknowledged the value of networks. For example, Al-Debei & Avison (2010) argue in their summary and review of IS research on business models that the success of a firm depends to some extent on the firm's relationships with other actors. However, networks as part of a business model are typically conceptualized and considered as value networks, that is, firm-level cooperation of suppliers, customers, and partners (Hedman & Kalling, 2003; Al-Debei & Avison, 2010; Zott *et al.*, 2011) as opposed to the personal networks and network resources available to the founders (i.e., social capital of founders). While social capital certainly helps establish firm-to-firm relations, the two concepts are distinct. Social capital is an individual-level concept, while value networks are a firm-level concept. In addition, social capital precedes value networks in that founders, when they begin to create a startup, typically do so with a substantial amount of combined social capital (probably the firm's only resource at this point), while the value network of the startup is typically non-existent at the beginning. Our argument picks up from a suggestion in the dynamic capabilities literature that 'social capital and external ties that individual team members bring with them may constitute important endowments of the founding team' (Helfat & Peteraf, 2003, p. 1001) in early-stage startups.

In summary, we are interested in the factors that make early-stage Internet startups successful and how the founders develop their business models. We deduce from theory and the literature that social capital previously overlooked may yet be an important factor in that success. As empirical studies directly exploring the mechanisms and testing the impact of the suspected effect are not available, we conducted an empirical investigation in the context of early-stage Internet startups.

Research Design

We used a sequential mixed-method design (Teddlie & Tashakkori, 2009; Creswell & Plano Clark, 2011) in our research. Mixed-method research has been strongly advocated in IS because it allows researchers to gain a more complete understanding of complex technological, organizational, and social phenomena of interest in our discipline (Avison & Fitzgerald, 2012; Ågerfalk, 2013; Venkatesh *et al.*, 2013).

Some scholars take the broader perspective that mixed-method research (or, multi-method research) refers to any research investigation that employs more than one method (Mingers & Brocklesby, 1997; Mingers, 2001; 2003), while other scholars are referring specifically to the combination of qualitative and quantitative research methods (e.g., Venkatesh *et al.*, 2013).[1] Our research design is mixed-method according to both perspectives.

We followed the most common type of mixed-method investigation, a sequential design (Creswell & Plano Clark, 2011).[2] First, we used interviews with Internet startup founders (i.e., expert informant interviews) to explore the issues qualitatively. These interviews were conducted in an open-ended fashion and we paid close attention to the interpretations and sense-making of the experts. The aim of this first part was to achieve a rich, context-aware exploration of the phenomena of interest – business model development and the role of social networks for early-stage Internet startups. Second, we used social network analysis to test quantitatively the central hypothesis of this paper, that is, that there is a strong, dominant, positive effect of social networks of founders on their startups' success. The integration of expert interviews and social network analysis has been singled out as one of the most fruitful applications of mixed-method research (Venkatesh *et al.*, 2013).

We describe both parts of our research – and how we integrated those two parts – in detail below.

Expert Informant Interviews

We conducted a series of semi-structured and open-ended interviews (Myers, 2013) with founders of Internet startups (see Appendix A for details on the participants). The interviews were based on an interview guideline we adapted over time to account for emerging insights or additional questions (see Appendix B for interview guidelines). We used a purposeful sampling strategy (Shadish *et al.*, 2001) in selecting and approaching the interviewees. Specifically, we

---

[1] All authors cited in this paragraph suggest that mixed-method research is valid (and valuable) – despite claims of the 'incommensurability' of the research philosophies justifying in particular research methods. The incommensurability claim (in relation to mixed-method research) has been disputed both by pragmatists and by critical realists on philosophical grounds. One core counter argument put forward against such a claim is that research methods provide *additional* and *complementary* insights on complex phenomena. Indeed, we found applying two different lenses (two methods) valuable for generating richer insights in relation to our research aim. For a more detailed discussion of the possibility and validity of mixed-method research consult, for example: Mingers & Brocklesby (1997), Mingers (2001) and Venkatesh *et al.* (2013).

[2] Note that sequential design refers to the logical design of the study. Some interviews (or follow-ups) were conducted after the quantitative analysis. In that sense, the research process was hermeneutic and iterative in nature (rather than one-way linear).

wanted to stay within specific boundaries (consumer Internet startups in or shortly after early stage) while maximizing variety within these boundaries (e.g., experience, location, business model type). To secure the preferred interview partners, we used the professional ties of members of the research team with startup organizations and the startup community.

The interviews took place via phone, video conferencing, or in person in 2013 and 2014. We transcribed the interviews in full to allow for coding analysis. We also included natural, secondary data such as press releases and publicly available third-party interviews (with other founders) for our analysis (Silverman, 2011). We used the software NVivo (Version 10) for coding. The conduct and analysis of the interviews were done iteratively and we adopted our interview guide over time to account for emerging insights (e.g., new aspects that seemed relevant).

For the coding and analysis process, we used the approach suggested by Lichtman (2013).[3] First, one author provided the initial coding of the data. In the initial coding process, we used both *in vivo* codes (i.e., codes that use terminology of the interviewees) as well as extant theory (i.e., terminology from the academic literature) to inform our naming and attribution of codes (Hsieh & Shannon, 2005). We then merged identical codes and resolved issues with different understanding of the meaning of codes in the research team through review, discussion, and clarification. Next, we aggregated codes to categories (higher-level constructs that describe several codes on a more abstract level). Again, we resolved different understanding of categories and interpretations through review, discussion, and clarification. We refined coding and categories over time by removing or merging unnecessary/redundant codes and categories. After developing categories, several researchers re-read all founder interviews to identify relationships between (e.g., structures in terms of cause-and-effect or timely order) and within categories (i.e., identified sub-categories). We then organized the categories into main concepts with respect to the main argument (i.e., linked to the research aim). The research team discussed the final categories, sub-categories, and their links, and resolved remaining ambivalences. This process was not linear. We conducted further interviews and iteratively

[3] Lichtman's process is in turn closely aligned with the coding techniques used in thematic analysis (e.g., Ezzy, 2002) or in the grounded theory method (e.g., Charmaz, 2014).

followed with the process of coding and abstracting (generating and relating categories) until changes in our understanding were minimal, and we concluded the investigation having reached theoretical saturation.

We used the qualitative findings to inform the quantitative part of the study. Specifically, the qualitative investigation allowed us to develop propositions and operationalize two of them into a hypothesis that could be meaningfully tested using (natural) quantitative data.

Social Network Analysis

For our quantitative data analysis, we collected different, complementary types of quantitative data. The data collection for this part of the study began in 2013 and was further complemented in 2014. We used several publicly available databases and sources to compile a unique dataset that encompassed 70 U.S. consumer Internet startups and their 145 founders. As the first step, we collected information from the CrunchBase (CrunchBase, 2014), AngelList (AngelList, 2013), and U.S. Securities and Exchange Commission (SEC, 2013) databases. These databases provided us with information about the business models of the startups and their funding information. We complemented this data with information available from the startups themselves (e.g., on their websites or in press releases). Second, we gathered publically available data from a large business online social networking site that provided indications of the founders' human capital (e.g., educational background or experience) and social capital (e.g., previous employers and experience – commonly held to be good indicators of social capital). The previous employers' data from the social networking site created a network of 29,419 companies. Note that we aggregated at the level of the startup in the case of co-founders.

We performed social network analysis (SNA) of the previous employers' network data to test for a correlation between the social network of founders and the success of their startups. SNA can determine properties of networks and nodes in these networks (Wasserman & Faust, 1994). Centrality, a common SNA measure, measures how central and well connected a node (here: a company) is. Eigenvector centrality, one particular SNA measure of centrality (Bonacich, 1972), assigns relative scores to all nodes in the network based on a recursive principle that connections to high-scoring nodes contribute more to the score of the node in question than connections to low-scoring nodes. Because we were looking at startups in a network of companies, we used Eigenvector centrality, which is considered an excellent SNA measure of social capital (Borgatti *et al.*, 1998). We chose not to use the directed weighted Eigenvector centrality because we wanted to include all previous employers – increasing diversity of inbound flows – rather than concentrate on a few high-intensity turnover relationships. Thus, we calculated the directed unweighted scaled Eigenvector centrality for each startup node. We used the software Gephi (Bastian *et al.*, 2009) for the SNA analysis.

Finally, we performed additional, more conventional statistical tests on the other data to test for correlation between a variety of other factors and the success of the startups. Depending on the nature of the underlying data, we either used a 2-sided Fisher's exact test or a regular 2-sided *t*-test (Hair *et al.*, 2010).

The findings of this analysis of the quantitative data provided further evidence to support the propositions developed earlier. Below, we first report the findings from the expert informant interviews (qualitative part of the study) before moving on to the findings from the social network analysis (quantitative part of the study). Further below, we then provide an integrated discussion of these findings.

Expert Informant Interviews: Analysis And Findings

All of our interviewed founders referred to Internet startup business models as being inherently dynamic. They considered the nature of startup business models to be substantially different from the business models of established companies. Most founders distinguished between the 'vision' (core idea), which was seen as stable, and the business model (operationalization), which was seen as volatile. This core idea was typically abstracted from specific details such as the value proposition or target customers, but it drove and influenced the development of a concrete business model.

Founder F explained:

'The core idea, for us, was always the same. What did change over time was everything else'.

The founders considered it precisely the purpose of an early-stage startup to 'define and refine' the business model. Specifically the context of Internet startups is an ideal environment to develop business models dynamically because required investments (e.g., development costs) are low and different configurations of the business model can in most cases be directly tested and iterated with potential customers.

The founders participating in the study were quite explicit that in early-stage Internet startups most elements of the business model need to be considered inherently dynamic. They attributed different degrees of dynamism to different elements of the emerging business model. A few considered single elements of their business model as being relatively stable, such as the intended monetization model. However, even in these cases the founders were not completely sure about the long-term stability of these elements either as a whole (e.g., moving from a business-to-business to a business-to-consumer business model) or at least in their adjustment over time (e.g., the exact price point inside a freemium monetization plan).

Based on these observations, we propose:

<u>Proposition 1</u>: The business models of early-stage Internet startups tend to be dynamic.

The founders themselves are the most stable element of the business model – or the startup if founders are considered external to the business model. The founders are the fixed point around which the business model evolves through multiple iterations.

An investor explained why he considers the business model less important than the founder team for his investment decisions:

'If bad comes to worse, can we reuse the founders for another [business] idea?' (Investor A)

Similar views regarding the pivotal role of the founders in early-stage Internet startups are held by leading investors such as Ron Conway and Chris Dixon (see Clark, 2011) or Dave McClure (see Patterson & Arnold, 2010). As the famous venture capitalist William Draper III concluded:

'You have to make sure that you have the right entrepreneur team. Nothing is more important. In fact, nothing is even a close second'. (see Lewis, 2011).

The founders conceptualized themselves as the creators or the enablers of the business model. They found themselves inherently central to the startup in its early stage. Summarizing the above, we propose:

Proposition 2: The founding team is mainly stable and plays a pivotal role for business model development in this early stage.

Rather than having a clear strategy in mind for their business models, the founders typically reported following trial-and-error, minimum-viable-product, and A-B-testing approaches[4] for their business model development. They further noted that the ability to adapt the business model quickly is more important than a clear strategy in the early stages of an Internet startup. Strategy was considered an abstract, high-level concept more relevant in later stages, after the business model was more stable. Founder G noted:

'Keeping agile. I would say being on my toes and changing directions; assessing things on the go is more important than sticking to a strategy which you are not very clear about'.

Based on these observations, we propose:

Proposition 3: The founders typically follow an agile trial-and-error approach in developing and adapting their business model.

What further stood out in the interviews was the high importance founders attributed to their social networks, especially professional social networks. All founders had leveraged in some way their social networks. Nearly all founders interviewed found it very important or even the most important factor for the success of their startup.

Founder J said his 'social network has a huge impact and it is amazing how it is working for me'.

[4] These approaches are short-cycle, iterative approaches to changing and testing the assumptions underlying a particular business model (e.g., Ries, 2011). The 'lean' and 'agile' approaches of business model development aim to reduce the number of spectacular large-scale earlier Internet venture failures often associated with the burst of the 'dot com bubble'.

Similarly, founder F said ‘my network is ... *the* success factor, to be honest’.

Founder E saw the social network as a resource of his startup: ‘The network is the most essential part of my company ...’

There were, however, exceptions to this general perception. Two of the 17 founders that participated in the study did not rely on their networks as centrally as the others. One stated that he used his social network only for emotional support, but that it was secondary for the business of the startup. Another preferred building strong personal relationships with a few people rather than leveraging a broad set of contacts.

Asked why their social networks were so important, the founders reported a broad range of benefits. More than half mentioned reputational benefits they gained through their networks. Having previously founded a successful startup or having worked at major consulting firms helped secure funding for the business model. Founder E explained his success in securing (Angel/Seed) funding as follows:

‘It would have never been possible to raise this much money in this short time, and having this much product without having said elements already. All of the investors are looking at my track record and everybody who knows me’.

All Asian founders corroborated the importance of status flows (i.e. flows of legitimacy, power, and recognition) through their social network and underscored the high importance of these reputational benefits for their startup. Conversely, for the Western founders, these were less central and they generally attached less importance to status flows.

Almost all founders reported information benefits from their social networks. For example, the founders reported having received relevant information and news, new ideas for their business model, and general feedback through their social networks. They also used their social networks as the key resource for advice on, and learning the skills necessary for dealing with, business, technology, or legal issues in their business model development.

Social networks were seen as providing important access to potential employees and external contractors (e.g., designers, lawyers, HR firms).

Networks were also central to many startups with respect to co-founders: they connected either specifically through searching for co-founders or by virtue of having previously worked or studied together. Some also found customers and partner companies through their social networks, and some gained access to office space and equipment.

Finally, one important benefit mentioned consistently in the interviews was that networks provided access to investors and hence funding. Nearly all the founders we interviewed received introductions to their future investors through contacts with which they already had a connection.

Investor A explained the importance of personal introductions through social networks:

'If they [founders] come to us [for funding] over network [through introductions], they need some kind of network. So, yes, the stuff we are looking at has a certain network ... They can use [a network] to get the right angels on board and the right one, two, three customers'.

The importance of social introductions and relationships for funding is generally accepted in practice (Nivi, 2008; Wertz, 2013). For example, Chris Wand, Managing Director at Foundry Group (a U.S.-based early-stage venture capital firm), explains:

'VCs are generally bombarded by requests for meetings, so a warm introduction helps an entrepreneur's request float to the top of the list'. (as quoted by Nivi, 2008)

In summary, we found in the data that the founders' social networks provide them with information, resource, and status benefits. For instance, they receive feedback on their ideas and concepts, gain partners and customers, get access to required resources, and raise funding through their networks. All of this helps them develop their business model and establish a successful startup.

<u>Proposition 4:</u> High social capital (i.e., being well connected) of the founders in early-stage Internet startups has a positive impact on the success of their startups.[5]

Based on the observations concerning the importance of founders' social networks, we spent considerable time in the interviews probing more deeply into the nature and origins of these social networks.

The founders reported that the relevant social networks included friends and family, in some cases formalized relationships (e.g., mentors), and most often professional contacts (especially from prior work, such as former co-workers). Such social relations can also lead to 'getting introduced' and 'establishing critical connections'.

Within the professional social network, founders found former co-workers as well as current or former investors and mentors to be the most helpful types of contact. Former co-workers (including bosses, mentors, or friends from previous workplaces) helped primarily by providing access to specific skills (e.g., legal advice, technical expertise) or through introductions to business-relevant contacts (e.g., investors, media). Typically, such transactions were social in nature and did not involve a monetary component. Friends and family were mentioned primarily as sources of emotional and motivational support.

All founders reported that they began working on their startups with substantial existing social networks based on their previous careers (especially employment and education). The key origin for their professional social networks was the founder's career history. Founder F explains this important relationship:

'I'm a lawyer. So, at the beginning, the patent lawyers, the corporation lawyers, and all these, they didn't charge a dime for us. ... When [co-founder's name] started talking about this to insurance companies, he always went directly to the board of directors because he did projects with them back then [in the co-founder's previous position]'.

[5] The aggregated benefit of being central to one or several social networks is typically referred to as 'social capital' in the literature. Some of the founders also used this term. For these reasons, we use the term social capital in the proposition.

In summary, the founders we interviewed stated that they benefitted most from professional contacts and that they actively managed their social networks. The most critical source for their social networks in the early stage stemmed from their career history, with some friends and family as well as investors and mentors also forming part of the social networks relevant for their business model development. We thus conclude with a final proposition:

Proposition 5: The founders' professional contacts stemming from their career history are the most important source of social capital.

Social Network Analysis: Analysis And Findings

We used propositions 4 and 5 to develop a testable hypothesis suitable for a larger-scale test (Whetten, 1989). To operationalize and measure 'social capital (i.e., being well connected) of the founders in early-stage Internet startups, we created the social networks of all founders, aggregated them per startup, and calculated the Eigenvector centrality as detailed in the method section and Appendix C. As noted above, if people are nodes in a network, then the Eigenvector centrality provides an excellent measure of their social capital (Borgatti *et al.*, 1998).

To operationalize and measure 'success of their startups', we considered a binary variable: 1 if the startup successfully secured Series A funding (considered the 'successful group') and 0 if the startup was unable to secure Series A funding (considered the 'unsuccessful group'). As noted above, Series A funding is widely considered the central and most important performance indicator for early-stage startups (Burton *et al.*, 2002; Davila *et al.*, 2003; Baum & Silverman, 2004).

Hypothesis: The combined social capital of the founders (as measured by their startup's Eigenvector centrality) has a significant positive impact on the success of their early-stage Internet startup (as measured by obtaining Series A funding).

We performed a t-test to assess the statistical significance of the group difference between the successful group and the unsuccessful group of startups. The result of the t-test confirmed that the means of the two groups are significantly different (t = -3.283, p = 0.002). Levene's test for equality of variances was not significant (p = 0.606), which indicates equal variances across both groups. Thus, the hypothesis was supported by the data at the 0.01 significance level.

Category	Measure	Successful startups	Unsuccessful startups	Significance of difference	Test type
Social capital	Founders' network centrality	21 (30%)	49 (70%)	0.002***	**b**
Human capital factors	Average number of (co-) Founders	2.238	2.000	0.270	**b**
	VC experience	3 (14%)	4 (8%)	0.421	**a**
	Consultant experience	6 (28%)	17 (34%)	0.783	**a**
	Prior founder experience	10 (47%)	34 (69%)	0.108	**a**
	Average years of education	4.786	3.699	0.074*	**b**
	Average years of work experience	9.572	9.446	0.936	**b**
Entrepreneurial self-efficacy	Management	19 (90%)	40 (81%)	0.485	**a**
	Financial	15 (71%)	36 (73%)	1.000	**a**
	Risk	16 (76%)	38 (77%)	1.000	**a**
	Innovation	11 (52%)	30 (61%)	0.599	**a**
	Marketing	13 (61%)	29 (59%)	1.000	**a**
Computer efficacy	General IT	15 (71%)	31 (63%)	0.590	**a**
	Web 2.0	9 (42%)	24 (48%)	0.795	**a**
	Design	2 (9%)	10 (20%)	0.325	**a**

a = Fisher's exact test; b = *t*-test (2-sided); significant at * α = 0.100; ** α =0.050, * α =0.010**

Table 1. Results from Tests of Association (N=70, Showing p-Values and Significance Levels)

In addition to testing the hypothesis (i.e., the impact of founders' social networks), we also controlled for some other factors that could potentially affect funding success. These include typical human capital factors such as education and work experience (Lazear, 2004; Lambert & Davidson, 2013), as well as specific entrepreneurial (see Chen *et al.*, 1998) and technological skills (see Compeau & Higgins, 1995) required in the context of Internet startups.

We found that founders' experience and specific skills were all statistically insignificant as antecedents of startup success, and that only years of education were moderately significant at the 0.10 level. The results indicate that human capital has a lower impact on startup success than social capital.

Discussion

The analysis of the qualitative data revealed that the business models of early-stage Internet startups are typically very volatile and in permanent flux. Despite that the founders attributed different degrees of dynamism to different elements of the emerging business model, even the relatively stable elements would or could change, or were at least adjusted over time. Thus, a business model in this context cannot be conceptualized meaningfully as an ontological static 'map' of a fixed structure already existing in reality. Rather, a business model in the early-stage phase is better understood as a conceptual structure for rapidly changing components that describe 'the rationale of how an organization creates, delivers, and captures value while delivering products or services to customers' (Osterwalder *et al.*, 2010, p. 14). The founders continually adapt and change the dynamic elements in an on-going search for a repeatable and scalable business model (see Blank & Dorf, 2012). In short, the business model is developed in this stage and changes significantly and rapidly. This finding adds empirical evidence to earlier research that considers business models as being inherently dynamic (Hedman & Kalling, 2003; MacInnes, 2005; Osterwalder *et al.*, 2005; Al-Debei & Avison, 2010; Trimi & Berbegal-Mirabent, 2012).

It was surprising that, as it turned out, the founders did not follow a clear strategy in developing their business model; rather, the business model evolved around a 'core idea'. This finding is in contrast to other research that considers an organization's strategic orientation as being the result of deliberate choices (e.g., Grover & Saeed, 2004).

We found that the founders' social networks – especially their professional network connections – are critical for the success of the startup (and hence the business model that is enacted through the startup). This finding supports prior research that suggests social networks provide information, resource, and status flows. To this literature, we add how these network benefits help in actually developing the business model.

As introduced before, we refer to Al-Debei and Avison (2010) for a more granular view on business models and their underlying dimensions – value proposition, value architecture, value network, and value finance. Information flows include the flow of information between individuals across their strong and weak ties (Friedkin, 1982), for instance competitive insights (Harrigan, 1986), but also knowledge transfers (Chrisman & McMullan, 2004). In our study, the founders mentioned having received both – new ideas and feedback for their business model as well as latest technological, business, and legal knowledge. Obtaining such knowledge is especially relevant at the early stage of a startup and has a significant impact on its long-term survival (Chrisman & McMullan, 2004). Consequently, information flows are relevant for all four business model dimensions; for example, constant feedback and new ideas help in shaping the value proposition as well as the revenue structure and pricing method (both part of the value finance dimension).

Resource flows include access to critical resources and transfer of assets such as money, equipment, and technology (Madhavan *et al.*, 1998; Gnyawali & Madhavan, 2001). In the early stage, the startup typically comprises the founder team (or a single founder) as its only resource. Our interview partners mentioned access to potential employees (or prospective co-founders), external contractors (e.g., developers, lawyers, etc.), customers or business partners, and investors through their networks as being critically important. These resource flows primarily help in further building and establishing the value architecture, that is, the organization's assets, resources, and core competencies. Furthermore, the access to customers and business partners help achieve product-market fit and hence an appropriate value proposition that creates value for customers.

Status flows are 'flows of legitimacy, power, and recognition from high-status firms to lower-status firms' (see Padgett & Ansell, 1993; Gnyawali & Madhavan, 2001, p. 432). In the early-stage startup context, our interviewees confirmed that they gained trust and credibility from having a proven track record or having worked for high-status firms, consulting companies, or relevant business partners. Thus, status flows play the role of 'door openers' that facilitate access to other actors who would then provide information and resource flows.

In summary, we conclude that the founders' social networks play a pivotal role for business model development because of the information, resource, and status benefits these networks provide.

Our quantitative data supports the hypothesis that centrality of founders in their professional social networks (stemming from founders' career histories) has a significant positive influence on the success of their startups (measured as Series A funding success). Conversely, typical human capital factors such as previous work experience or specific skills were not significant. This finding is in line with previous research that showed that social capital has significantly greater impacts than human capital in explaining startup success (Stam *et al.*, 2014). Yet, we were able to show these effects in the unique context of early-stage Internet startups, which generally poses great challenges for researchers due to limited data availability (George & Bock, 2011).

Drawing meta-inferences – that is, the integration of findings across the qualitative and quantitative studies – is a critical and essential aspect of mixed-methods research (Venkatesh *et al.*, 2013). Here, both our research streams complemented each other well. The first part, the qualitative investigation through expert interviews, revealed that the founders themselves see their professional social networks as critically important for the development of the business model. To complement and expand on this rather subjective perception of the founders, we performed a subsequent quantitative statistical analysis of archival data, which was (implicitly) based on the more objective evaluation of investors. The quantitative analysis confirmed that better-connected founders have higher chances to obtain required follow-up funding for their startups. Combining the findings from both research streams, we conclude that startups with better-connected founders are more successful because the founders' professional social networks provide the required means (i.e., information, resource, and status benefits) to develop the business model in the early stage of the startup.

What are the implications from these findings for future conceptualization of and research on business models and startups? First, our findings provide new and unique insights into how young digital organizations dynamically develop their business models. Thus, we address a research gap identified by Al-Debei & Avison (2010) as well as George & Bock (2011) by adding how these business models actually emerge in practice to earlier research that primarily focused on what comprises a business model. Specifically, this paper helps us understand business models in the empirical context of early-stage Internet startups as unstable and in flux. In this context, business models cannot be treated as a static fact but must be conceptualized as a dynamic structure that is emergent and constantly adapted over time.

Conversely, we found the founders to be a relatively stable component (of the startup and/or the business model). They are a central factor in explaining early-stage Internet startup success. This adds to previous research on the link between business model and firm performance that looked at more mature firms (Andries & Debackere, 2007; Zott & Amit, 2007). Further, these findings imply that an integration and synthesis of IS research on business models (typically focused on the organizational value proposition model) with entrepreneurial studies (typically focused on the person of the founder) will be fruitful to develop our knowledge of business models (recently also suggested by Veit *et al.*, 2014).

Second, the findings provide further evidence in a new context for the claim that external networks are a critically important resource for firms (on organizational level; e.g., Gnyawali & Madhavan, 2001; Canina *et al.*, 2012). In particular, we show how early-stage Internet startups leverage the information, resource, and status benefits their founders' networks provide to develop their business models. Notably, we found empirical evidence that the founders' social ties serve as antecedents to value network connections (e.g., to customers and suppliers). This finding certainly warrants further investigation of the link between individual-level and organizational-level networks in the context of business model and entrepreneurship research.

Finally, methodologically, this study is among the first to integrate qualitative methods and SNA methods in an IS research paper. In their recent review of mixed-method IS research published in top IS journals (AIS senior scholar basket, which includes the *Information Systems Journal*), Venkatesh *et al.* (2013) were unable to identify such an example, yet strongly and specifically advocated for this particular combination. As such, this paper may both provide an insightful account for researchers interested in the context of business models and IS entrepreneurship and also provide an example of how qualitative methods and SNA can be meaningfully integrated in IS research in general. We found the two methods mutually informing, and found it encouraging overall that similar results through different methodological tools result in the same empirical finding.

For Internet entrepreneurs and prospective founders, the findings suggest the critical importance of actively building professional social networks before creating a startup, and then nurturing and using these social networks extensively to develop a business model that adds value for the customer (rather than operating under the alternative assumption that 'it just takes the right idea'). Often – especially in the pay-it-forward culture of the Internet startup scene (Blank, 2011; Ready, 2012) – such networks provide ‘free’ (in monetary terms) critical resources for the business model/startups. ‘Don’t ask for money, ask for advice’ (Calacanis, 2011) is a common mantra in the Internet startup scene. Increasing the social capital of a founder team, according to our findings, is a crucial mechanism to improve the likelihood of success of the startup. In terms of actionable advice to prospective founders, this would suggest getting a person with high social capital on board (e.g., a person with long professional experience, a well-connected angel investor, or an experienced serial entrepreneur). In that sense, both our participants as well as other practitioners suggest that the relationship between founder team and investor is (for both parties) not primarily a financial but a social arrangement (e.g., Dupree, 2012; Espinal, 2013; Kerpen, 2013).

For angel investors and other early investors in Internet startups – a group that could rapidly expand from an elite group to ‘everybody’ in the near future through crowdfunding and crowd investing/equity (Belleflamme *et al.*, 2014; Mollick, 2014) – the findings have implications for how to evaluate Internet startups. Investors should look beyond the business model in isolation and the ‘human capital’ competencies of the founders. They should study the networks founders bring to the table. That is, a startup with a good business idea will likely not succeed in securing required later-stage (Series A) funding if the founders lack the professional social capital required to develop a successful business model.

While our mixed-methods approach allowed us to provide rich insights for the reader, our study is limited in several ways. The first limitation is due to its empirical context. The Internet startup sector is a large and important sector (especially for IS researchers). However, this sector also has some unique characteristics that could lead to the findings not holding in other empirical contexts (or at least requiring additional tests if used for theorizing beyond the defined boundaries). Hence, we cannot and do not claim universal ‘generalizability’ of findings (see further Lee & Baskerville, 2003). Still, we believe the findings do provide a starting point for further research and theorizing in other contexts as well.

Second, while we had sufficient samples to reach theoretical saturation in the key informant interviews and reach statistically valid results in the statistical test, the sample size is somewhat limited (17 founders in the qualitative part + 70 startups and their 145 founders in the quantitative study). While these numbers compare to other sample sizes used in SNA (Stam *et al.*, 2014), future work could enlarge the size of both empirical data considered. In the quantitative part, we studied exclusively Western (mainly U.S.-based) startups, which means we were not able to control for regional or cultural impacts. However, the results of the qualitative study, which included Chinese startups[6], indicate little differences between early-stage Internet startups in the Western world and in China. One difference we identified, however, is the stronger importance of status flow in China. This relates to the Chinese concept of 'guanxi', which describes the deeply embedded cultural importance of personal networks of influence (see further Davies *et al.*, 1995; Tsang, 1998; Park & Luo, 2001). Cross-industry or cross-cultural extension of the data set will provide further insights in this regard.

Third, we used 'natural' data from public sources. While using natural data prevents some of the problems and biases of survey-based research (e.g., natural data is generated through actual business performance in practice, not through the efforts of the researchers), it is restricted to data that are directly available and that typically are incomplete or may have other issues. For example, it appears that most Internet startups are listed in CrunchBase, but we had no control over startups not listed in CrunchBase, which makes any claim of a 'complete sample' problematic.

Finally, it should be clearly noted that we do not present (nor claim to present) a full theory of business models. Rather, we have developed five propositions and tested one resulting hypothesis. This is certainly useful for future theorizing, but will require further empirical studies of other contexts and the consideration of other factors to further develop such a theory. This study may provide a building block (Weick, 1995) for future theory developments.

Conclusion

Business models of early-stage Internet startups are highly dynamic and continually changing; in fact, the business model is only developed in this stage. Startup success is not a predetermined function of opportunity identification and the 'right idea'; rather, such success is the result of a very dynamic process of business model iteration and validation. The founders themselves play a crucial role in early-stage Internet startup success (considered in this study

[6] We thank the reviewers of this paper for pushing us to extend the scope of the qualitative study towards more geographic and cultural diversity.

as the ability to achieve Series A funding), especially through their social capital. Using a mixed-method approach, we found strong support that startups with better-connected founders are more successful because the founders' professional social networks provide the required means (i.e., information, resource, and status benefits) to develop the business model in the early stage of the startup. We show this link based on 17 expert informant interviews with founders around the globe and the application of social network analysis to a unique sample of 70 startups and the professional networks of their 145 founders. While further empirical and theoretical work on business models is needed, we believe these findings provide an important piece in better understanding business model development and firm success in the context of early-stage Internet startups.

References

Afuah, A. & Tucci, C. L. (2001) *Internet Business Models and Strategies: Text and Cases*. McGraw-Hill, Boston, MA, USA.

Ågerfalk, P. J. (2013) Embracing diversity through mixed methods research. *European Journal of Information Systems*, **22**, 251–256.

Al-Debei, M. M. & Avison, D. (2010) Developing a unified framework of the business model concept. *European Journal of Information Systems*, **19**, 359–376.

Andries, P. & Debackere, K. (2007) Adaptation and performance in new businesses: understanding the moderating effects of independence and industry. *Small Business Economics*, **29**, 81–99.

AngelList (2013) AngelList - Home. [WWW document]. URL http://angel.co (accessed 5 August 2014).

Ardichvili, A., Cardozo, R. & Ray, S. (2003) A theory of entrepreneurial opportunity identification and development. *Journal of Business Venturing*, **18**, 105–123.

Avison, D. & Fitzgerald, G. (2012) Reflections and opinions on 25 years with the ISJ. *Information Systems Journal*, **22**, 179–193.

Barney, J. B. (1986) Strategic factor markets: expectations, luck, and business strategy. *Management Science*, **32**, 1231–1241.

Bastian, M., Heymann, S. & Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. *Proceedings of the 3rd Int'l AAAI Conference on Weblogs and Social Media (ICWSM),* 17-20 May, San Jose, California.

Baum, J. A. C. & Silverman, B. S. (2004) Picking winners or building them? alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing*, **19**, 411–436.

Belleflamme, P., Lambert, T. & Schwienbacher, A. (2014) Crowdfunding: tapping the right crowd. *Journal of Business Venturing*, **29**, 585–609.

Bhagavatula, S., Elfring, T., van Tilburg, A. & van de Bunt, Gerhard G. (2010) How social and human capital influence opportunity recognition and resource mobilization in India's handloom industry. *Journal of Business Venturing*, **25**, 245–260.

Blank, S. (2005) The Four Steps to the Epiphany: Successful Strategies for Products that Win. Cafepress.com, San Mateo, CA, USA.

Blank, S. (2011) The pay-it-forward culture - Steve Blank on wordpress.com. [WWW document]. URL http://steveblank.com/2011/09/15/the-pay-it-forward-culture/ (accessed 5 August 2014).

Blank, S. & Dorf, B. (2012) The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company. K & S Ranch Press, Pescadero, CA, USA.

Bonacich, P. (1972) Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, **2**, 113–120.

Borgatti, S. P. & Foster, P. C. (2003) The network paradigm in organizational research: a review and typology. *Journal of Management*, **29**, 991–1013.

Borgatti, S. P., Jones, C. & Everett, M. G. (1998) Network measures of social capital. *Connections*, **21**, 27–36.

Burkhart, T., Krumeich, J., Werth, D. & Loos, P. (2011) Analyzing the business model concept - a comprehensive classification of literature. *Proceedings of the 32nd International Conference on Information Systems,* 4-7 December, Shanghai, China.

Burton, M. D., Sørensen, J. B. & Beckman, C. M. (2002) Coming from good stock: career histories and new venture formation. In: *Research in the Sociology of Organizations*, Lounsbury, M., Ventresca, M. J. (eds), pp. 229–262. Emerald Group Publishing Ltd., Bingley, UK.

Calacanis, J. (2011) Startups - Cal Tech talk. [WWW document]. URL https://www.youtube.com/watch?v=ehDJB2K5Q6I (accessed 5 August 2014).

Canina, L., Palacios, D. & Devece, C. (2012) Management theories linking individual and organizational level analysis in entrepreneurship research. *International Entrepreneurship and Management Journal*, **8**, 271-284.

Charmaz, K. (2014) *Constructing Grounded Theory*. SAGE Publications, Thousand Oaks, CA, USA.

Chen, C. C., Greene, P. G. & Crick, A. (1998) Does Entrepreneurial Self-Efficacy Distinguish Entrepreneurs from Managers? *Journal of Business Venturing*, **13**, 295–316.

Chesbrough, H. (2010) Business model innovation: opportunities and barriers. *Long Range Planning*, **43**, 354–363.

Chesbrough, H. & Rosenbloom, R. S. (2002) The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, **11**, 529–555.

Chrisman, J. J. & McMullan, W. E. (2004) Outsider assistance as a knowledge resource for new venture survival. *Journal of Small Business Management*, **42**, 229–244.

Clark, B. (2011) Startup success: how 7 top angel investors do business. [WWW document]. URL http://mashable.com/2011/10/26/how-angels-invest/ (accessed 5 August 2014).

Cohen, W. M. & Levinthal, D. A. (1990) Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, **35**, 128–152.

Compeau, D. R. & Higgins, C. A. (1995) Computer self-efficacy: development of a measure and initial test. *MIS Quarterly*, **19**, 189–211.

Creswell, J. W. & Plano Clark, V. L. (2011) *Designing and Conducting Mixed Methods Research*. SAGE Publications, Los Angeles, CA, USA.

CrunchBase (2014) CrunchBase - Home. [WWW document]. URL http://www.crunchbase.com/ (accessed 5 August 2014).

Davies, H., Leung, Thomas K. P, Luk, Sherriff T. K & Wong, Y.-h. (1995) The benefits of “guanxi”: the value of relationships in developing the Chinese market. *Industrial Marketing Management*, **24**, 207–214.

Davila, A., Foster, G. & Gupta, M. (2003) Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, **18**, 689–708.

Del Giudice, M. & Straub, D. (2011) IT and entrepreneurism: an on-again, off-again love affair or a marriage? *MIS Quarterly*, **35**, iii–vii.

Dupree, J. R. (2012) G-Startup – investing in startups is like marriage, say VCs at GMIC SV. [WWW document]. URL http://www.mobisights.com/2012/10/20/g-startup-investing-startups-marriage-vcs/7670 (accessed 5 August 2014).

Espinal, C. E. (2013) What tier is your investor (or what to look for in an investor)? [WWW document]. URL http://www.seedcamp.com/resources/what-tier-is-your-investor-or-what-to-look-for-in-an-investor (accessed 5 August 2014).

Ezzy, D. (2002) Qualitative Analysis: Practice and Innovation. Routledge, London, UK.

Friedkin, N. E. (1982) Information flow through strong and weak ties in intraorganizational social networks. *Social Networks*, **3**, 273–285.

Fruchterman, T. M. J. & Reingold, E. M. (1991) Graph drawing by force-directed placement. *Software: Practice and experience*, **21**, 1129–1164.

George, G. & Bock, A. J. (2011) The business model in practice and its implications for entrepreneurship research. *Entrepreneurship Theory and Practice*, **35**, 83–111.

Gnyawali, D. R. & Madhavan, R. (2001) Cooperative networks and competitive dynamics: a structural embeddedness perspective. *The Academy of Management Review*, **26**, 431–445.

Granovetter, M. S. (1985) Economic action and social structure: the problem of embeddedness. *The American Journal of Sociology*, **91**, 481–510.

Grover, V. & Saeed, K. A. (2004) Strategic orientation and performance of Internet-based businesses. *Information Systems Journal*, **14**, 23–42.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010) *Multivariate Data Analysis - A Global Perspective*. Pearson, Upper Saddle River, NJ, USA.

Harrigan, K. R. (1986) *Managing for Joint Venture Success*. Lexington Books, Lexington, MA, USA.

Hedman, J. & Kalling, T. (2003) The business model concept: theoretical underpinnings and empirical illustrations. *European Journal of Information Systems*, **12**, 49–59.

Helfat, C. E. & Peteraf, M. A. (2003) The dynamic resource-based view: capability lifecycles. *Strategic Management Journal*, **24**, 997–1010.

Hsieh, H.-F. & Shannon, S. E. (2005) Three approaches to qualitative content analysis. *Qualitative Health Research*, **15**, 1277–1288.

Kerpen, D. (2013) 3 ways angel investors can help you (a lot). [WWW document]. URL http://www.inc.com/dave-kerpen/angel-seed-funding-success.html (accessed 5 August 2014).

Kollmann, T. (2006) What is e-Entrepreneurship? fundamentals of company founding in the net economy. *International Journal of Technology Management*, **33**, 322–340.

Krumeich, J., Burkhart, T., Werth, D. & Loos, P. (2012) Towards a component-based description of business models: a state-of-the-art analysis. *Proceedings of the 17th Americas Conference on Information Systems,* August 9-12, Seattle, Washington.

Lambert, S. C. & Davidson, R. A. (2013) Applications of the business model in studies of enterprise success, innovation and classification: an analysis of empirical research from 1996 to 2010. *European Management Journal*, **31**, 668–681.

Lazear, E. P. (2004) Balanced skills and entrepreneurship. *American Economic Review*, **94**, 208–211.

Lee, A. S. & Baskerville, R. L. (2003) Generalizing generalizability in information systems research. *Information Systems Research*, **14**, 221–243.

Lewis, N. (2011) Business angels; its not about the product! [WWW document]. URL http://www.ibusinessangel.com/2011/06/business-angels-its-not-about-the-product/ (accessed 5 August 2014).

Lichtman, M. (2013) *Qualitative Research in Education: A User's Guide*. SAGE Publications, Los Angeles, CA, USA.

MacInnes, I. (2005) Dynamic business model framework for emerging technologies. *International Journal of Services Technology and Management*, **6**, 3–19.

Madhavan, R., Koka, B. R. & Prescott, J. E. (1998) Networks in transition: how industry events (re)shape interfirm relationships. *Strategic Management Journal*, **19**, 439–459.

McGrath, R. G. (2010) Business models: a discovery driven approach. *Long Range Planning*, **43**, 247–261.

McKnight, L. W., Vaaler, P. M. & Katz, R. L. (2002) *Creative Destruction: Business Survival Strategies in the Global Internet Economy*. MIT Press, Cambridge, MA, USA.

Mingers, J. (2001) Combining IS research methods: towards a pluralist methodology. *Information Systems Research*, **12**, 240.

Mingers, J. (2003) The paucity of multimethod research: a review of the information systems literature. *Information Systems Journal*, **13**, 233–249.

Mingers, J. & Brocklesby, J. (1997) Multimethodology: towards a framework for mixing methodologies. *Omega*, **25**, 489–509.

Mollick, E. (2014) The dynamics of crowdfunding: an exploratory study. *Journal of Business Venturing*, **29**, 1–16.

Morris, M., Schindehutte, M. & Allen, J. (2005) The entrepreneur's business model: toward a unified perspective. *Journal of Business Research*, **58**, 726–735.

Morris, M. H., Schindehutte, M., Richardson, M. & Allen, J. (2006) Is the business model a useful strategic concept? conceptual, theoretical, and empirical insights. *Journal of Small Business Strategy*, **17**, 27–50.

Muzyka, D., Birley, S. & Leleux, B. (1996) Trade-offs in the investment decisions of European venture capitalists. *Journal of Business Venturing*, **11**, 273–287.

Myers, M. D. (2013) *Qualitative Research in Business & Management*. SAGE Publications, London, UK.

Nahapiet, J. & Ghoshal, S. (1998) Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, **23**, 242–266.

Nivi, B. (2008) Who makes the best introductions to investors? [WWW document]. URL http://venturehacks.com/articles/introductions (accessed 5 August 2014).

Nonaka, I. (1994) A dynamic theory of organizational knowledge creation. *Organization Science*, **5**, 14–37.

NVCA (2013) National Venture Capital Association yearbook 2013. [WWW document]. URL http://www.nvca.org/index.php?option=com_content&view=article&id=257&Itemid=103 (accessed 5 August 2014).

Osterwalder, A., Pigneur, Y., Bernarda, G. & Smith, A. (2014) Value Proposition Design: How to Create Products and Services Customers Want. Wiley.

Osterwalder, A., Pigneur, Y. & Clark, T. (2010) Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Wiley, Hoboken, NJ, USA.

Osterwalder, A., Pigneur, Y. & Tucci, C. L. (2005) Clarifying business models: origins, present, and future of the concept. *Communications of the Association for Information Systems*, **16**.

Padgett, J. F. & Ansell, C. K. (1993) Robust action and the rise of the medici, 1400-1434. *American Journal of Sociology*, **98**, 1259–1319.

Park, S. H. & Luo, Y. (2001) Guanxi and organizational dynamics: organizational networking in Chinese firms. *Strategic Management Journal*, **22**, 455–477.

Patterson, I. & Arnold, C. (2010) Seed capital from angel investors: Dave McClure, founder, 500 Startups (part 7). [WWW document]. URL http://www.sramanamitra.com/2010/10/18/seed-capital-from-angel-investors-dave-mcclure-the-founder-500-startups-part-7/ (accessed 5 August 2014).

Poole, G. (2001) The riddle of the abandoned shopping cart. *The Industry Standard*, **December–January**, 77–82.

Porter, M. E. (1985) *Competitive Advantage*. Free Press, New York, NY, USA.

Ready, K. (2012) Paying it forward: Silicon Valley's open secret to success. [WWW document]. URL http://www.forbes.com/sites/kevinready/2012/08/23/paying-it-forward-silicon-valleys-open-secret-to-success/ (accessed 5 August 2014).

Ries, E. (2011) The Lean Startup: How Constant Innovation Creates Radically Successful Businesses. Portfolio Penguin, London, UK.

SEC (2013) Filings & forms. [WWW document]. URL https://www.sec.gov/edgar.shtml (accessed 5 August 2014).

Serarois-Tarrés, C., Padilla-Meléndez, A. & del Aguila-Obra, A. R. (2006) The influence of entrepreneur characteristics on the success of pure dot.com firms. *International Journal of Technology Management*, **33**, 373–388.

Shadish, W. R., Cook, T. D. & Campbe, D. T. (2001) Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Wadsworth Publishing.

Shane, S. & Venkataraman, S. (2000) The promise of entrepreneurship as a field of research. *The Academy of Management Review*, **25**, 217–226.

Shirky, C. (2008) Here Comes Everybody: The Power of Organizing Without Organizations. Penguin, London, UK.

Silverman, D. (2011) Interpreting Qualitative Data: A Guide to the Principles of Qualitative Research. SAGE Publications, Los Angeles, CA, USA.

Stam, W., Arzlanian, S. & Elfring, T. (2014) Social capital of entrepreneurs and small firm performance: a meta-analysis of contextual and methodological moderators. *Journal of Business Venturing*, **29**, 152–173.

Taylor, C. (2012) Study: more than 1,000 seed-funded startups are destined to be 'orphaned' at the Series A stage. [WWW document]. URL http://techcrunch.com/2012/12/19/study-more-than-1000-seed-funded-startups-are-destined-to-be-orphaned-at-the-series-a-stage/ (accessed 5 August 2014).

Teddlie, C. & Tashakkori, A. (2009) Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences. SAGE Publications, Los Angeles, CA, USA.

Teece, D. J. (2010) Business models, business strategy and innovation. *Long Range Planning*, **43**, 172–194.

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

Trimi, S. & Berbegal-Mirabent, J. (2012) Business model innovation in entrepreneurship. *International Entrepreneurship and Management Journal*, **8**, 449–465.

Tsang, E. W. K. (1998) Can guanxi be a source of sustained competitive advantage for doing business in China? *The Academy of Management Executive (1993-2005)*, **12**, 64–73.

Unger, J. M., Rauch, A., Frese, M. & Rosenbusch, N. (2011) Human capital and entrepreneurial success: a meta-analytical review. *Journal of Business Venturing*, **26**, 341–358.

Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J. M., Loos, P. & Spann Martin (2014) Business models – an information systems research agenda. *Business & Information Systems Engineering*, **6**, 45–53.

Venkatesh, V., Brown, S. A. & Bala, H. (2013) Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, **37**, 21–54.

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

Weick, K. E. (1995) What theory is not, theorizing is. *Administrative Science Quarterly*, **40**, 385–390.

Wernerfelt, B. (1984) A resource-based view of the firm. *Strategic Management Journal*, **5**, 171–180.

Wertz, B. (2013) The best way to approach investors: four tips. [WWW document]. URL http://versiononeventures.com/best-way-approach-investors-four-tips/ (accessed 5 August 2014).

Whetten, D. A. (1989) What constitutes a theoretical contribution? *Academy of Management Review*, **14**, 490–495.

Zahra, S. A. & George, G. (2002) Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review*, **27**, 185–203.

Zhao, H., Seibert, S. E. & Lumpkin, G. T. (2010) The relationship of personality to entrepreneurial intentions and performance: a meta-analytic review. *Journal of Management*, **36**, 381–404.

Zott, C. & Amit, R. (2007) Business model design and the performance of entrepreneurial firms. *Organization Science*, **18**, 181–199.

Zott, C., Amit, R. & Massa, L. (2011) The business model: recent developments and future research. *Journal of Management*, **37**, 1019–1042.

Appendix A: Overview of Interview Partners

Name	Founded in	Funding	Country	Business context
Investor A	-	-	Germany	-
Founder A	2010	Seed	USA	Customer service
Founder B	2010	Seed	Germany	Mobile apps
Founder C	2011	Angel (2 rounds)	Germany	Web business
Founder D	2008	Series A	Germany	Social network
Founder E	2013	Seed	Germany	Social advertisement
Founder F	2011	Seed (2 rounds)	Germany	Automotive mobile apps
Founder G	2012	Bootstrapped	USA	Business intelligence
Founder H	2010	Angel	Germany	News
Founder I	2012	Seed	Germany	Social media
Founder J	2013	Seed	USA	E-Commerce
Founder K	2012	Bootstrapped	USA	Web business
Founder L	2013	Seed	China	Web business
Founder M	2013	Accelerator	China	Mobile business
Founder N	2013	Angel	China	Web business
Founder O	2013	Seed	China	Mobile games
Founder P	2012	Seed	China	Web business

Appendix B: Interview Guideline (Key Informants Study)

Interview topic	Question
Founding event	Opening question: Please describe how (and when) the new/last business was founded?
	When was your new/last business founded?
	How many co-founders (if any) were there when the startup was founded?
Funding	Opening question: How has the business been funded over time?
	Did you receive angel investment? When? Optional: How much? From whom?
	Did you receive seed funding? When? Optional: How much? From whom?
	Did you receive Series A funding? When? Optional: How much? From whom?
	Did you receive further funding other than angel, seed, and Series A funding? Optional: How much? From whom?
Team constellation (optional)	Have the founding team members changed over time? Did founders leave the initial team or did you add additional team members as co-founders (not as regular employees)?
	Is there a 'lead entrepreneur' – a single founder who had the business idea in the first place and then searched for suitable co-founders?

	Are all founders equal – for example, gain-risk-share, personal investments into the business, etc.?
Business model (underlying dimensions and their relative importance)	In your opinion, what are the business model characteristics of Internet startups? What makes these business models distinctive? How do these business models differ from other industries, such as traditional technology, biotech, and so on? How do you see the development of an idea over time? Does the idea often stay the same or does it change? How do you see the development of the business model of your startup? Was it static or did it change dynamically, evolving over time? (Note that some see only the finance model as 'business' model.) Please provide some more details about how your business model evolved over time? Which business model elements were present (fixed) in the beginning and stable over time, if any? Would you say that your entire business model changed over time, or that only some underlying elements changed? Is there a well-thought-out strategy behind how you evolve your business model, or do you follow more of a trial-and error approach?
Business model (underlying dimensions and their relative importance)	Are there some cornerstones (core beliefs) in the strategic orientation of your business model – for example, a strong focus on innovation vs. operational efficiency? Do these cornerstones change over time? How to you do the terms 'business idea', 'strategy', and 'business model' relate to each other? If you need to balance strategic thinking with the ability to adapt quickly, which would you say is more important for a founder in an early-stage startup? Would you consider the founders (yourself) to be an inherent element of the business model (e.g., a critical resource) or are the founders not part of the business model but instead its creators and enablers? Why? Generally speaking, what do you see as the role of the founder in the early stage of a startup with regards to the business model?
Founder's network activities	What role does your social environment and social network play – that is, your relationships with friends, former colleagues, and other contacts (social capital)? Do you actively manage and grow your social networks? How? Support from strong ties: Did you receive any support from your spouse/life partner, parents, friends, and/or relatives during your founding activity? What kind of support? Support from weak ties: Did you receive any support from business partners, acquaintances, former employers, and former co-workers? What kind of support? Looking back, who are the six most important people who supported you in building up your business?

Founders' social capital, connectedness, and network benefits	Opening question: What is more important in founding a successful startup: human capital (the things you can do by yourself) or social capital (the people you know that can do things for you)? Why? Have you specifically observed/benefitted from your network, contacts, etc.? Resource flow: e.g., former colleagues who joined your startup as co-founders or regular employees, etc. Information flow: e.g., 'insider' knowledge; early access to relevant information for your business (industry trends, new products to be launched, etc.) Status flow: e.g., increased credibility in the business context; advance of trust; etc.

Appendix C: Data Collection and Correction for Quantitative Study

We used two primary data sources for our research effort: CrunchBase, a free public database that provides structured profiles on technology startups, including detailed funding information (CrunchBase, 2014); and a large business online social networking site for more detailed information on founders as well as turnover information between companies. In addition to these two databases, we referred to the startup platform AngelList (AngelList, 2013) and to the 'About Us' pages of the startups we included in the study to gather additional information on the number of founders. Finally, we also used press mentions and the public database of the U.S. Securities and Exchange Commission (SEC, 2013) to validate the funding information from CrunchBase. Table 2 is an overview of the data sources and the type of data collected.

Data source	Type of data	Purpose of data
CrunchBase	Structured profiles of technology startups, including detailed funding information	Sample selection
Large business online social network	More detailed information on founders, especially their employment history, as well as turnover between companies.	Supplement, validation
AngelList	Information on the number of founders.	Supplement
U.S. Securities and Exchange Commission	Funding information	Validation

Table 2. Overview of data sources and types

First, we needed to identify a suitable sample of early-stage Internet startups for our analyses. We selected all startups in the CrunchBase database that belonged to the category ‘Consumer Web’, were U.S.-based, and received seed funding in 2011. This resulted in a relatively homogenous first set of 103 startups and their 188 founders. Next, we referred to the business online social network, where we searched manually for our founders’ profiles. We were able to find profiles for 94 per cent; the other 6 per cent did not have a profile. Unfortunately, not all the profiles included complete resumes, so we had to exclude those founders and their related startups from our subsequent analyses. Thus, we ended up with a final set of 70 startups and their 145 founders. The founders we removed from the data set showed no patterns other than the fact that they had not completely filled out their profiles. As we were only interested in their job histories, we assumed they were missing completely at random (MCAR).

We coded most of the information from the founder’s profiles into an Excel workbook, including demographics, educational background, and specifically prior employers (see Appendix F below for details). For the startups, we used the funding information to differentiate successful and unsuccessful startups using a dichotomous (dummy) variable based on whether they had received Series A funding.

Second, we created our industry network. The business online social network provided insightful statistics on turnover between companies. For every company, there was a statistic of the top 5 companies from which people came and where they went, based on its vast database of people and their individual career histories. The basic idea was to follow those inbound and outbound ties between companies and eventually capture a network of turnover relationships. However, this direct approach would have had the disadvantage of collecting only the largest (top 5) turnover relationships while neglecting all others. We circumvented this limitation by beginning the process not with a single company but rather with a large set of companies derived from CrunchBase. We followed a 3-step ‘snowball’-approach: First, we downloaded all companies from CrunchBase (set of 87,257) and tried to match them with companies present on the business online social network, using its API. This resulted in 23,659 companies. In Step 2, we accessed the company profile page on the business network of every company in the CrunchBase set, and followed all inbound and outbound turnover connections of these companies. As the third step, we accessed all new companies that resulted from step 2 and again followed their turnover relationships. Theoretically, this should result in 10 more companies for every company. Practically, the number of new companies converged quickly. We halted the process after repeating step 3 twice, resulting in a set of 29,419 companies. Some 95 per cent of these companies were connected in one giant component, while the remaining 5 per cent were distributed across small, disconnected islands of a maximum of 8 companies each. We removed these otherwise unconnected companies, as they would not influence the calculation of our network attributes. This resulted in a final industry network of 27,857 companies that were connected to each other by 85,767 turnover relationships.

Third, we integrated our 70 startups into the industry network. Similar to the process described above, we used the turnover information of the founders to connect startups with existing companies in the industry network. For instance, if a founder had worked previously for IBM, Yahoo, and SAP, ties were added from the founder's startup to those three companies. We applied social network analysis methods to investigate and differentiate startups based on their relative position in the industry network (see Appendix F).

Appendix D: Transformation of Eigenvector Centrality

Before transformation, the Eigenvector Centrality (EC) variable showed significant deviations for skewness and kurtosis (see Table 3), as depicted in the left Q-Q plot in Figure 1 below. In addition, the modified Kolmogorov-Smirnov test of normality shows that the data significantly deviates from a normal distribution.

	Skewness		**Kurtosis**		**Test of Normality**[a]	
Variable	Statistic	z value	Statistic	z value	Statistic	Significant.
Eigenvector Centrality (EC)	2.011	7.007	4.105	7.253	.239	.000
EC after LN transformation	-.048	-.167	-.958	-1.693	.083	.200[b]

[a] Results are shown for the modified Kolmogorov-Smirnov test (Lilliefors significance correction). In addition, the Shapiro-Wilk test was also not significant (p=.065).

[b] Lower bound of true significance

Table 3. Test of Normality

After transformation using the natural logarithm (LN), the data were found to be normally distributed, as confirmed by the test of normality and the second Q-Q plot (see Hair *et al.*, 2010 for a discussion of variable transformations).

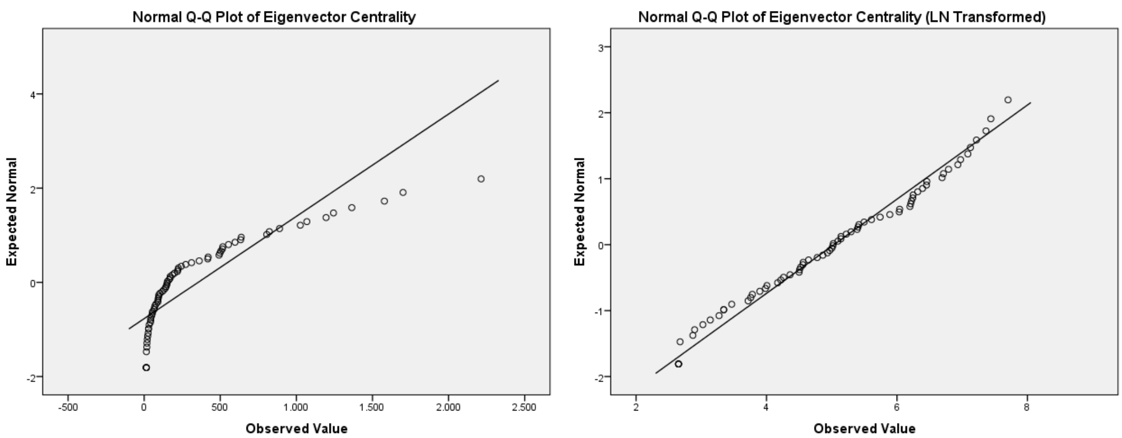
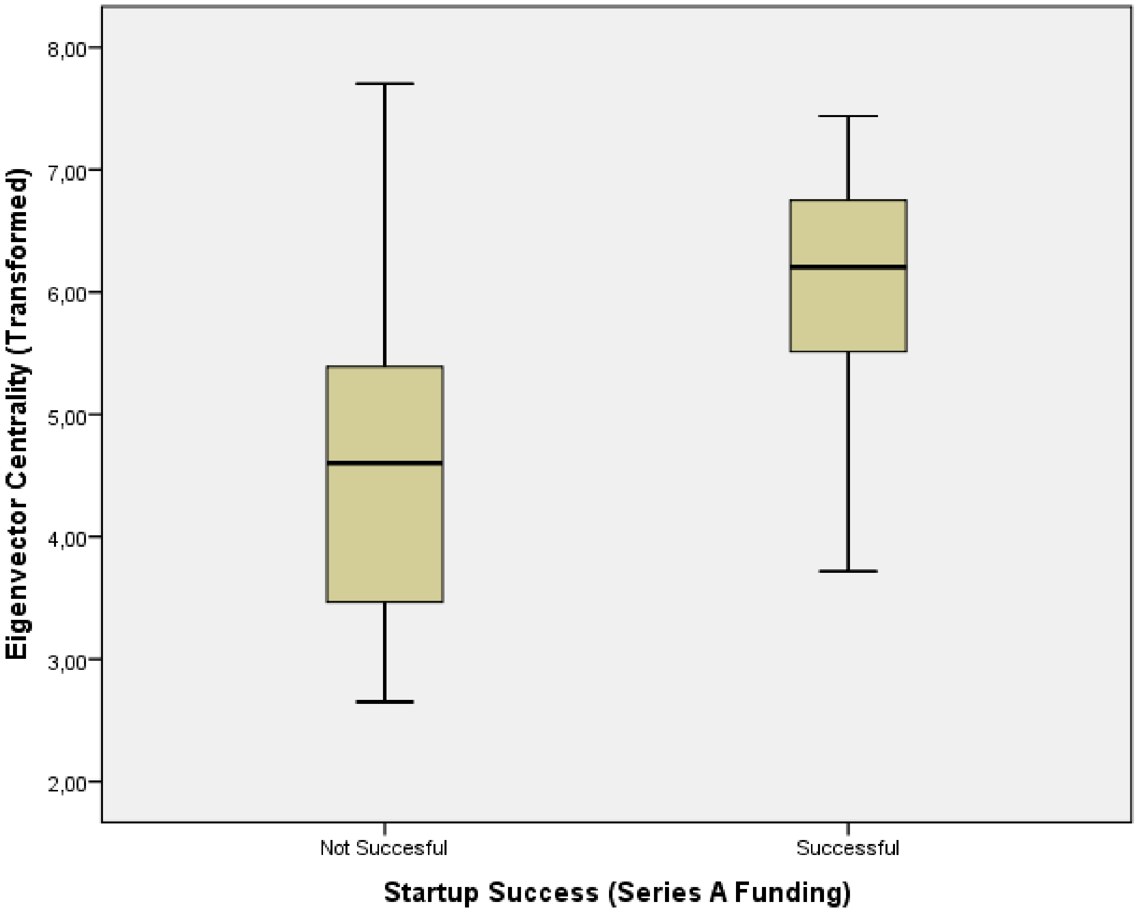


Figure 1. Q-Q Plot Before and After Data Transformation

Appendix E: Eigenvector Centrality of Successful vs. Unsuccessful Startup Groups



Appendix F: Industry Context (Network Graph)

The Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991) has been applied to calculate the spatial positions of nodes in the turnover industry network. Fruchterman-Reingold is a force-directed algorithm that considers a repulsive force between any two nodes. This force causes well-connected nodes of a network to attract each other, and places nodes with high centrality in more central positions. Consequently, the structural layout of the graph reflects the proximity and cohesion of nodes in the turnover network (see Figure 2 below). Black circles (+) represent the successful and white circles (-) the unsuccessful startups in our sample. Black squares represent the most central companies in the network, including IBM, Microsoft, Google, and SAP.

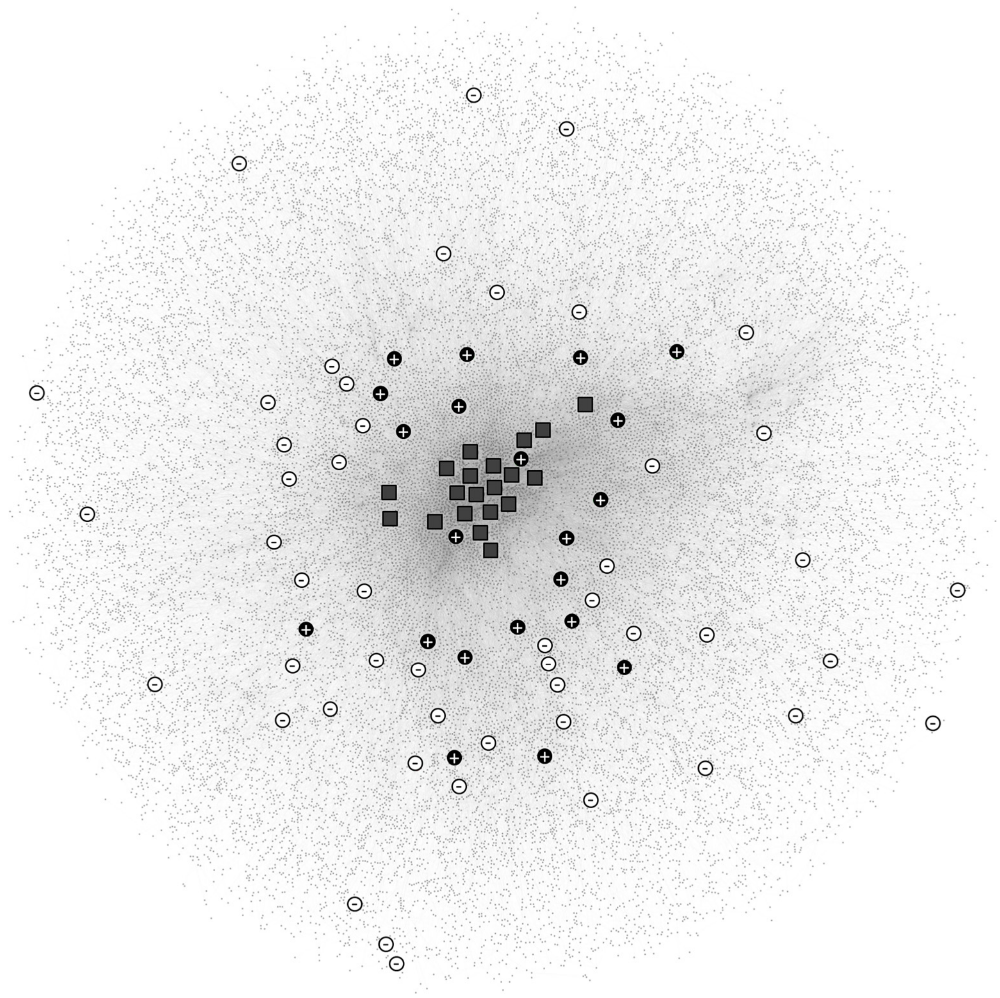


Figure 2. Industry network

Paper IV

Bibliographic data	Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Fischbach, K., Schlagwein, D., & Schoder, D. (2014). *Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors.* In: Proceedings of the Academy of Management Annual Meeting. Philadelphia, PA, https://doi.org/10.5465/ambpp.2014.15984abstract
Status of publication	Published.
Publication type	Conference paper
Version & Copyright notice	Accepted version After a paper is accepted for publication in the Annual Meeting Proceedings, the Academy of Management grants to the author(s) the following rights: • The right to make and distribute copies of all or part of the paper for the Author(s) own use in teaching, research or for internal distribution within the institution/company that employs the Author(s) provided that such copies are not resold. • The right to use and publish, after release of the Annual Meeting Proceedings, all or part of the material from the paper in any original or derivative work.
Previous versions	n/a

Getting Boundary Conditions Right: Towards a Classification of the Information Economy Sectors

Just as most people are not WEIRD, the assumption of industry uniformity may not be true in all cases. Several reviews showed that IS research does not take industry seriously enough. Neglecting industry context can have a severe effect on research results by underspecifying theory or by leading to general explanations that do not hold in other contexts. This paper examines the so called “Information Economy”, an industrial context comprising of the ICT, Content and Media, and Internet sectors. We analyze a unique, very large data set that contains employee mobility data of 27,387 organizations. We derive some interesting descriptive statistics that help to differentiate the Information Economy sectors. In addition, with the application of a clustering algorithm, we derive industrial clusters on our data. Our analysis reveals that the OECD's conceptualization of the Information Economy reaches its limits when it comes to more granular sub-sectors within the industry, at which point it no longer seems appropriate from a social/cultural industry perspective. Our study contributes to ongoing discussions around generalizability and boundaries of research results, as well as to the still small body of industry research in the IS field. These findings have important implications for future research and practice.

Keywords. IS research, generalizability, industry context, turnover, employee mobility, Information Economy, ICT sector, Internet sector, network analysis, industry classification.

Introduction

Just as most people are not WEIRD (Western, Educated, Industrialized, Rich, and Democratic) (Henrich, Heine and Norenzayan, 2010a, 2010b), the assumption of industry uniformity may not be true in all cases (Chiasson and Davidson, 2005). As Henrich et al. (2010b) have shown at the individual level, bad sampling leads to inappropriate validity, generalization, and boundary claims. Chiasson and Davidson (2005) have highlighted the same problem and have demanded that industry be taken more seriously in IS research. Their review of research articles published in *MIS Quarterly* (MISQ) and *Information Systems Research* (ISR) between 1997 and 2004 found that most (58%) they analyzed did not identify the industry in which the study was conducted. They noted: "[I]ndustry provides an important contextual 'space' to build new IS theory and to evaluate the boundaries of existing IS theory" (Chiasson and Davidson, 2005). A review a few years later by Seddon and Scheepers (2012) of all articles published in *MISQ* and *ISR* in 2007 and 2008 found that only 24 percent of the empirical papers discussed the boundary conditions that applied to their results.

The issue of generalizability of research findings has gained some attention by IS researchers in recent years (esp. Lee and Baskerville, 2003; King and Jun He, 2005; Tsang and Williams, 2012; Lee and Baskerville, 2012). "A research generalization is the researcher's act of arguing, by induction, that there is a reasonable expectation that a knowledge claim already believed to be true in one or more settings is also true in other clearly defined settings" (Seddon and Scheepers, 2012). Generalization requires that researchers clearly define their target population (i.e., the population from which they draw a sample and to which they want to generalize their results) and discuss potential boundary conditions that might apply to their findings (Seddon and Scheepers, 2012). Neglecting industry context can have a serious effect on research results by underspecifying theory or by leading to general explanations that do not hold in other contexts (Chiasson and Davidson, 2005). "Identifying individuals, groups, and organizations that are members of an industry is an empirical, as well as a conceptual, issue" (Chiasson and Davidson, 2005). While high-level industry categories such as "manufacturing" or "service sector" are more appropriate for meta-analyses, more granular government classification schemes may be useful to define industry boundaries (Chiasson and Davidson, 2005). For instance, Xue, Ray, and Gu

(2011) use the *North American Industry Classification System* (NAICS) to measure business unrelatedness between headquarters and business units as a function of differences in industry codes. The statistical tests in these sorts of studies, which rely on classification systems such as International Standard Industrial Classification (ISIC) or NAICS, are, however, only valid to the degree the underlying classification system is valid. Industry classifications such as ISIC define industries by similar production processes, similar products, or similar behavior in financial markets. These approaches are imperfect, and companies get lumped into sectors as though they do only one thing (Karmarkar and Apte, 2007).

The challenge of coming up with a narrow definition of IS-based industries is complicated by the pervasiveness of Information and Communication Technology (ICT) in other industries and the so-called "convergence" of technology and industries (e.g., Garcia-Murillo and MacInnes, 2003). The Organization for Economic Co-operation and Development (OECD) sees the ICT sector as part of the larger construct "information society/economy," which "... encompasses the widely agreed elements of ICT supply, ICT demand, ICT infrastructure, ICT products and content" (OECD, 2011). Along with "Content and Media" and "ICT demand, use, and e-commerce by businesses," it is part of the umbrella term "Information Economy" (see OECD, 2011). While the Information Economy may be defined in other ways (Karmarkar and Apte, 2007), we employ the OECD definition of the Information Economy not as synonymous with the "Knowledge Economy" but as an umbrella economic sector comprising other sectors, and hence our use of the term Information Economy throughout the remainder of this paper should be read as referring to this economic sector.

The terms "Information and Communication Technology (ICT) sector" or "ICT industry" have been used interchangeably in past studies, often leading to confusion about the actual scope and potential boundaries (cf. Chiasson and Davidson, 2005). Only recently, some efforts have been made to define the ICT and Content and Media sectors (OECD, 2011) based on the ISIC (United Nations, 2008). The existence of an Internet (or e-business) sector also was acknowledged by the United Nations working groups; however, as the Internet permeates all industries, a clear definition based on a subset of ISIC codes was not feasible (United Nations, 2008).

According to Crowston and Myers (2004) industries can be analyzed from three different yet complementary research perspectives: economic, institutional, and social/cultural. While the economic perspective is most commonly taken in IS research, the other two offer several advantages. For instance, the social/cultural perspective focuses on social relationships, networks, and structure that can be built by patterns of interaction among people and organizations within an industry (Crowston and Myers, 2004). It includes employee mobility, recently identified as an interesting point of research from a social network analysis perspective: Collet and Hedström (2013) show that tie formation in interorganizational networks generated by employee mobility are contingent upon the direction of past ties and that most of the knowledge exchange stemming from this mobility occurs on short sociometric distances. Employee mobility, furthermore, can be used to determine the existence of a cluster and the degree to which it functions as an industrial cluster. Other than certain case-based economic geographical studies (Martin and Sunley, 2003; Agrawal, Cockburn and McHale, 2006), no systematic attempts have been made to define industrial systems based on turnover flows (Eriksson and Lindgren, 2008). Initial findings from Eriksson and Lindgren (2008) promise a helpful systematic method to determine industry clusters based on employee mobility data.

The purpose of this study is to explore the Information Economy from a social/cultural industry perspective and evaluate the use of industry classifications to define and identify potential boundaries in the form of distinct industry sectors and sub-sectors. Our study addresses the following research question: How do industry boundaries differ when they are derived "top-down" from an industry classification or generated "bottom-up" based on employee mobility?

This research is based on a unique, very large dataset comprising aggregated employee mobility data from 27,387 companies within and outside the Information Economy. We apply descriptive statistics as well as a social network perspective, especially a clustering algorithm, to identify patterns and structures based on employee mobility.

The remainder of this paper is organized as follows. First, we provide some theoretical foundations. Next, we describe our research method and data, and then present our findings. We discuss the research results next, as well as our contributions, potential limitations, and implications of our research. We conclude with a brief summary and a research agenda.

The Information Economy and its Context in Industry Studies

Today, catchphrases such as "IT" (Information Technology), "ICT," "online media," and "the Internet" have become part of our daily vocabulary. Terms such as "ICT sector" are often used as if common and clear definition existed. However, what the ICT sector actually includes changes over time, varies between studies, and is often not clearly defined. For statistical and econometric models, such definitions should include a specific list of industries under the umbrella of each term. The challenge of coming up with a universal definition is further complicated by the pervasiveness of ICT in other industries and the so-called "convergence" of technology and industries (e.g., Garcia-Murillo and MacInnes, 2003). The use of the term ICT in both the narrow and wide senses adds to the general confusion. For instance, Fransman refers to ICT as comprising "... telecoms, IT, consumer electronics and Internet/media" (Fransman, 2010). Conversely, the Organization for Economic Co-operation and Development (OECD) sees the ICT sector as part of the larger construct "information society/economy," which "...encompasses the widely agreed elements of ICT supply, ICT demand, ICT infrastructure, ICT products and content" (OECD, 2011).

In this paper, we build on and extend the OECD (2011) definition for two reasons. First, there is a clear advantage to using the term ICT in a narrower sense, because it is more precise and avoids unnecessary confusion. Second, in recent years, there has been considerable work from different international groups to define the ICT sector, as well as a Content and Media sector (see OECD, 2011), based on the ISIC industry classification system developed by the United Nations Statistics Division (United Nations, 2008). Meanwhile those definitions have also been incorporated into the ISIC standard itself as alternative aggregations (United Nations, 2008).

Again, the OECD uses "Information Economy" as a superordinate term for an entire ecosystem. Two sectors have already been detailed and defined based on ISIC classifications: the ICT sector (OECD, 2011), and a Content and Media sector (OECD, 2011). In addition, there is a third sector in the Information Economy referred to as "ICT demand by businesses" and "ICT use and e-commerce by businesses." For practical reasons, we refer to the latter as the Internet sector, since there is no such clear-cut definition for the Internet sector in terms of ISIC classifications as there are for the other two sectors. This is because "[p]roduction units engaged in e-commerce will ... be found in any industry of ISIC" (United Nations, 2008) and thus cannot be directly mapped to a set of singled-out ISIC industries. The OECD defines e- commerce as "the sale or purchase of goods or services, conducted over computer networks by methods specifically designed for the purpose of receiving or placing of orders. The goods or services are ordered by those methods, but the payment and the ultimate delivery of the goods or services do not have to be conducted online. An e-commerce transaction can be between enterprises, households, individuals, governments, and other public or private organisations" (OECD, 2011).

Overall, there has been little research on the Information Economy as a whole. Most previous studies refer to the ICT industry but do not include exact industry or sector definitions. Thus, the following literature review includes studies that refer to the ICT industry or industries.

There is a paucity of research on industry in Information Systems (IS) research in general, and among the empirical articles in top-tier IS journals that have addressed industry, the high- tech industry – which comprises the ICT sector – accounted for only about 20 percent of the publications (Chiasson and Davidson, 2005). There are only a few more recent industry studies that look at the ICT ecosystem as a whole (e.g., Fransman, 2010; Arlanis and Ciriani, 2010; Adomavicius et al., 2007). In addition, little empirical research has been conducted on the industry level; most IS research has focused on the organizational and individual levels (Crowston and Myers, 2004).

The ICT industry has been studied as an ecosystem in which companies interact, innovate, and compete through convergence, bundling, and external growth strategies (Arlanis and Ciriani, 2010). This connected view of the industry at a company level is also employed in studies at a deeper technological level in which technology evolution is viewed as an ecosystem (Adomavicius et al., 2007). Koski and Siermo (2003) go further to study the ICT industry as a network industry and show how it differs from non-ICT industries in form of entry and exit dynamics of firms. Very few studies have looked at the ICT sector from a network perspective. Existing studies aim to show the network of alliances (Hallikas et al., 2008) or to shed a light on cooperation and competition in the ICT industry (Ritala, Hallikas and Sissonen, 2008).

Beyond certain economic geographical studies (Eriksson and Lindgren, 2008; Martin and Sunley, 2003; Agrawal, Cockburn and McHale, 2006), no systematic attempts have been made to define industry classifications based on turnover flows. Initial findings from Eriksson and Lindgren (2008) promise a helpful systematic method to determine industry clusters based on employee mobility data.

Research Methods and Data

In the following, we introduce our data sources and sampling strategies. Next, we describe the Social Network Analysis methods we applied. Finally, there is a brief outline of our data cleansing, validation, and coding efforts.

Data Sources and Sampling

Measuring the Information Economy is not an easy task, especially at the company-level. We use data from a leading worldwide professional social networking site that provides insightful statistics on turnover between companies. Companies as well as people have profiles on this site. Every company page provides statistics of the top-5 other companies from which its employees either came or went, derived from its vast database of people and their individual career histories. Following those inbound and outbound ties between companies would allow one to capture a network of turnover relationships; however, such a direct approach has the disadvantage that collects only the largest (top 5) turnover relationships and neglects all others.

We overcame this limitation by beginning not with a single company but with a large set of companies (the so-called seed) derived from CrunchBase, a free public database that provides structured profiles of technology companies (CrunchBase, 2013). We followed a 3-step "snowball"-approach. First, we downloaded all companies from CrunchBase (87,257 as of May 3, 2012) and tried to match them with companies present on the professional social networking site, using its API. This yielded a seed of 23,659 companies. We then accessed the company profile page of every company in the seed and extracted all information relevant for our research, including company size, type, industry, and year of founding. As the third step, we followed all inbound and outbound turnover connections of these companies. Theoretically, this would result in 10 more companies for every company in the best case, that is, without duplicates. Practically, the number of new companies converged quickly. We halted the process after repeating the second and third steps twice, ultimately yielding a set of 29,419 companies. Data extraction was completed in June 2012.

Social Network Analysis Methods

Social Network Analysis models actors and their interrelationships in networks of nodes and ties (also referred to as edges). Actors can be individuals, groups, companies, or even more abstract entities such as patents. Ties represent a certain type of relationship: they can be directed (e.g., seeks advice from) or undirected (e.g., know each other). In addition, relationships may be modeled as binary (exists or not) as opposed to signed, ordinal, or valued (e.g., the amount of trade between two countries). A network can be fully connected or decomposed into components that are connected internally, but not connected among each other (like islands in the sea) (Wasserman and Faust, 1994).

For our research, we used the above mentioned turnover data to generate a directed weighted network of companies. Weighted ties between any two companies reflect the sum of employee turnovers between these two companies. The direction of a tie represents the employee turnover from company A to company B.

It is our goal to exploit the network described above to identify and compare cohesive industry structures, that is, clusters comprised of similar companies. Accordingly, we perform a cluster analysis on the turnover network described above.

Cluster analysis, a well-established method in statistics, is “a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess” (Hair et al., 2010). Unlike in discriminant analysis, the clusters are not predefined; rather, the goal is to identify them (Hair et al., 2010). Identifying structures in networks (i.e. community detection) is challenging and has been studied intensely in graph theory (Fortunato, 2010). Community detection algorithms can reveal naturally emerging communities by exploiting a network’s topology. Many of those algorithms are limited. As demonstrated by (Lancichinetti and Fortunato, 2009), even well established methods often fail to produce valid results in large networks. Moreover, many algorithms are not designed for weighted, directed, or weighted and directed networks

(Lancichinetti and Fortunato, 2009). Consequently, if applied to such networks, they neglect a considerable amount of information carried by those and produce results that do not necessarily reflect the dynamics underlying a network's structure (e.g., knowledge exchange, employee mobility).

Rosvall and Bergstrom (2008) have introduced a community detection algorithm especially suited for large networks with weighted and directed ties; it has performed very well in several benchmarking studies (see Lancichinetti and Fortunato, 2009). The central idea behind the algorithm, which we further refer to as Infomap, is to find a good partition of a network by tracking the movements of a random walker surfing on its edges. Considering edge directions and weights, the walker reveals good communities by getting trapped in cohesive subgroups of nodes. Using greedy search and simulated annealing techniques, the method is capable of revealing multi-level hierarchies, which are represented by distinct groups (i.e. clusters) of nodes on several hierarchy levels (Rosvall and Bergstrom, 2011).

By applying the algorithm to the turnover network discussed above, we are able to identify a hierarchy of clusters based on turnover flows. Since the algorithm tries to maximize the internal cohesion of those clusters, we expect them to reflect adequately the dynamics of employee movements, as described by Collet and Hedström (2013).

Data Cleansing, Validation, and Coding

Most of the companies (95 percent) in our initial data set were connected with each other in one giant component. The remaining 5 percent were distributed across small, disconnected islands, each with no more than 8 companies. We removed these otherwise unconnected companies, resulting in a preliminary dataset of 27,857 companies. These companies were classified into 145 different industries. However, classification of industries on LinkedIn does not follow any established classification system. Thus, we used research team triangulation to match the LinkedIn industries to the respective ISIC classifications and Information Economy sectors (see the overview of classification systems above). In the course of our analyses, we verified this high-level classification multiple times for selected companies and corrected their respective industry codes if necessary. For instance, the IT service and consulting company Accenture classified itself in "Management Consulting," while in

ISIC terminology the company is part of the ICT sector. When in doubt, we looked up additional company information on Businessweek.com (Bloomberg, 2013).

Findings

In this section, we describe our research findings along two separate streams of analysis. First, we present some descriptive characteristics of the three Information Economy sectors. These have been derived by taking a top-down approach on our data based on the existing OECD sector definitions and their underlying ISIC industry classifications. Second, we present the results from our cluster analysis, where we applied a community detection algorithm (Infomap) to our employee mobility network data to identify distinct sub-sectors in a bottom-up approach.

Descriptive Characteristics of the Information Economy Sectors Based on OECD Definitions and ISIC Classification (Top-Down)

Here we briefly summarize the descriptive statistics derived from our data set.

Employee Mobility Across Industries. As noted above, our final dataset comprised 27,387 companies. Of those companies, 14,789 fall into the Information Economy category. Within this category, 10,918 companies belong to the ICT sector, 2,106 to the Content and Media sector, and 1,765 to the Internet sector. Other industries account for 13,068 companies. There are less companies in other industries due to our sampling approach that focused on the Information Economy, so to say these other companies are only the closest neighbors.

Out of the 1,056,982 relevant employee movements captured in our network, 81 percent occur within the Information Economy, 10 percent are inbound from other industries, and the remaining 9 percent outbound from the Information Economy to other industries.

Employee Mobility Across and Within the Information Economy Sectors. Now we are going to take a closer look at the employee mobility within and across the three sectors in the Information Economy. Table 1 below summarizes the result.

Insert Table 1

about here

The first number in each cell of Table 1 represents the share of the total amount of employee mobility in the Information Economy. Not surprisingly, the ICT sector dominates employee mobility as it comprises of most of the companies. Overall, the ratio between inbound and outbound employee mobility is balanced, with one exception: the Internet sector receives slightly more employees than it is losing to the other sectors (about 1.5:1).

As described earlier, every employee that moves from employer A to employer B creates a directed tie from A to B. On a sector level, we can sum these ties and count how many are within a sector (between companies in the same sector) or how many connect different sectors (ties between companies in different sectors). The result of this simple calculation would be biased, again, as there are many more companies in the ICT sector than in the other two sectors. We corrected for this bias by account for the theoretical distribution of ties[1]. The results are also presented in Table 1 above. The numbers in brackets represent the deviation of observed tie distribution from what would be expected from a uniform distribution of ties between companies. We find that there are more turnover relationships (ties) between companies within the same sector than expected (above zero values in the main diagonal). For example, in the ICT sector there are 19.91 percent more ties than would be in a uniform distribution. Conversely, the connectedness between any two sectors is lower than it would be in the case of a uniform distribution of ties (below zero

[1] In a set of n companies, there are $(n \times (n - 1))$ possible directed ties. Between two groups of n and m companies, there are $(n \times m)$ possible directed ties. In the total set of both groups of companies, there are $((n + m) \times (n + m - 1))$ possible directed ties.

numbers in all cells but the main diagonal). But there are differences. The ICT and Content and Media sectors are more separated (less connected) from each other than the Content and Media and Internet sectors among each other.

These results are also supported from a graphical network perspective. Figure 1 shows four different views of the same network graph. The first graphic on the left shows the Information Economy as a whole. The three other graphics highlight each of the three sectors.

Insert Figure 1

about here

Three observations can be made. First, the different sectors cluster into three distinct areas (compared to a random distribution), and there is more cohesion within the sectors than across. Second, the Content and Media sector is somewhat more separated from the two other sectors. Third, the Internet sector appears tightly integrated into the ICT sector.

Company Sizes in the Information Economy Sectors. There is a clear indication that the Internet sector comprises of much more small and middle-sized companies. While 40.4 percent of the companies in the Internet sector have 1-50 employees, there are only 26.1 percent in the ICT sector, and with 18.9 percent even less in the Content and Media sector. Conversely, there are more companies $>$ 5000 employees in the ICT (6.1%) and Content and Media sector (8.5%), than in the Internet sector (1.6%).

Company Types in the Information Economy Sectors. Most of the companies (73.4%) in the three sectors combined are privately held, while 23.1 percent are public companies. The Internet sector has above average privately held companies (80.3%) and well below average public companies (16.3%). Notably, 2.9 percent of the Content and Media sector companies are nonprofit organizations; a closer look reveals that these are public broadcasting companies such as the BBC, as well as other media companies such as Oxford University Press, the National Geographic Society, and the Associated Press.

Company Ages in the Information Economy Sectors. In the Internet sector, 62.2 percent of the companies are less than 9 years old, compared to only 29.2 percent in the Content and Media sector, and 35.1 percent in the ICT sector. The other end of the scale is dominated by companies from the Content and Media sector, where 21.8 percent are older than 50 years, compared to only 3.1 percent in the ICT sector, and 0.2 percent in the Internet sector. These numbers need to be treated with some caution, as only 65 percent of the companies in our dataset had the company age filled in.

Towards a Classification of Information Economy Sectors Based on Employee Mobility (Bottom-Up)

The results from the cluster analysis – after applying the community detection algorithm to our network – are presented in Table 2.

The two network diagrams in the first row of Table 2 show the direct results from the clustering. The size of each node is determined by the number of companies that are part of the cluster. The thickness of ties between any two clusters shows the weighted connectedness. The diagrams only show a subset of the total number of clusters, that is, the top 10. The diagram on the right upper corner shows a zoom into "Cluster 1," which is shown on the left side.

The spider charts in Table 2 each show the distribution of companies in a certain cluster along 7 ISIC classifications labeled as the letters "A" to "G" which are described in the legend at the bottom. For presentation purposes, the underlying percentage values have been logarithmized. Otherwise smaller percentage values would not have been visible in the charts. The headline in each cell, e.g., "Cluster 1.8", refers to the cluster numbers in the network diagrams. Cluster 1.8 is the 8th sub-cluster in Cluster 1 on the root level. Thus the three spider charts in the second row of Table 2 describe the cluster 1, 2, and 3 from the first diagram, and the remaining 10 smaller spider charts describe the first ten clusters below cluster number 1 on the root level, that is, all clusters in the network diagram on the right.

Overall, the algorithm identified 135 clusters on the first (root) level. However, the distribution has a long tail. The first 10 clusters account for 97.42 percent of the companies (see running totals in Table 3). The top 3 clusters show distinct patterns in the underlying classification, as can be seen in the first three spider charts. Cluster 1 primarily consists of companies in ICT-related classifications (letters E,F, and G), and there are some companies related to the Internet sector (letter A). However, there is almost no trace of companies belonging to the Content and Media sector (letters B, C, and D). On the contrary, Clusters 2 and 3 are almost purely characterized by the Content and Media sector. Surprisingly, in Cluster 3 there are also some companies related to the ICT sector (letter F). We investigated this further and found that this cluster also comprises companies at the interface with the Content and Media sector. For instance, there are companies that provide information services and access to content databases, as well as some computer software providers.

Zooming into Cluster 1, the algorithm identified a total of 1,235 sub-clusters. Again, there is a very long tail; the median of the cluster size is only 4. The top 10 sub-clusters still account for 32.36 percent of the companies in Cluster 1. As can be seen from the different patterns in the respective spider charts, each sub-cluster has a very unique "fingerprint." For instance, while both sub-clusters 1.1 and 1.5 cover letters F and G, sub-cluster 1.1 is dominated more by letter F (computer programming, consultancy, and related activities), and sub-cluster 1.5 by letter G (manufacture of computer, electronic, and optical products). Sub- cluster 1.5 also includes more companies that belong to letter E (telecommunications). Thus, one could summarize that sub-cluster 1.5 is more technology and manufacturing focused than sub-cluster 1.1.

We further investigated the top 10 clusters and sub-clusters to obtain a better understanding of their characteristics beyond letter codes and ISIC classifications. The results are presented in Table 3. For each cluster, we report a describing tag, a list of major companies belonging to the cluster, and the total number of companies of which the cluster is comprised. The last column of the table shows how much of a cluster's employee turnover flows within the cluster.

Insert Table 2

about here

Insert Table 3

about here

Discussion

In this section, we first discuss our research results. Next, we highlight our contributions and implications for research and practice, before we finally address the limitations of this study.

Research Results

From a top-down perspective of characteristics of the Information Economy sectors based on OECD definitions and ISIC classification, we find there is much greater turnover within the Information Economy than with industries outside. In addition, the inbound and outbound flow of employees is almost balanced. Taking a closer look at industries and companies that exchange employees with the Information Economy, we find there are considerable differences between the three sectors. Most interesting, the analysis of the top 10 sources for employees reveals some characteristics: the ICT sector receives employees from outsourcing partners and the military; the Content and Media sector primarily from financial companies; and the Internet sector from universities and consulting firms. This is indicative of the different maturity levels of these sectors. ICT products and services have increasingly become commodities in most industries, and thus outsourcing non-core activities is quite common. On the contrary, the Internet sector is still developing and, at the same time, quite innovative, not only attracts but also requires highly skilled and motivated (young) professionals directly from the universities or from professional services firms.

These differences between the sectors are found as well when looking at company size, type, and age. The Internet sector, for instance has much smaller and younger companies that are predominantly privately held. Conversely, the ICT and Content and Media sectors have many more companies with more than 10,000 employees and significantly more public companies. Again, this reflects the different maturity levels of those sectors.

Looking more closely at the Information Economy sectors, we find that the industry is not as homogeneous as one might think, at least from a turnover perspective. First, in absolute terms, there is much more turnover within each of these sectors than across. This results in distinct industry clusters that are also visible in a network graph. The Internet sector is closely related to the ICT sector, but still clearly distinguishable, while the Content and Media sector is more an adjacent neighbor of the other two sectors. These clusters exist not only on a sector level, but some unique sub-sectors can also be identified and located. The ICT sector itself consists of some local clusters, including network equipment, semiconductors, services (consulting and outsourcing), and software and platforms. The video game industry subsector plays an interesting role in connecting companies in the Content and Media sector with companies in the ICT sector. Another important finding is that we can identify companies that bridge several subsectors. Thomson Reuters and Electronic Arts are prominent examples. Thomson Reuters links four different subsectors. It is essential for multinational information and media firms to attract employees from different sectors to gain competitive advantages over its rivals.

Contributions and Implications

Our research contributes to ongoing discussions around generalizability and boundaries of research results, as well as to the still small body of industry research in the IS field. In line with Weick's (1995) argument, our paper is not theory, but an important building block for future theorizing. We find that the three sectors ICT, Content and Media, and Internet actually have very distinct characteristics in terms of employee mobility, as well as company size, type, and age. Thus, on a high level, we find empirical support for the OECD's conceptualization of the

Information Economy based on the ISIC industry classification. Further, our bottom-up cluster analysis reveals that the OECD's conceptualization of the Information Economy reaches its limits when it comes to more granular sub-sectors within the industry, at which point it no longer seems appropriate from a social/cultural industry. Finally, we provide an alternative approach to classify sectors and sub-sectors based on real life employee mobility data.

Our research has some implications for research and practice. First and foremost, our studies show that context actually matters a lot. The Information Economy is not a homogeneous industry, at least not from a turnover perspective. Rather, there are distinct differences between the sectors and even sub-sectors. Future research in these industries should take account of these findings and consider potential boundary conditions that might apply. This includes making it very explicit what the scope of an industry study actually is, e.g., referring to the ICT sector either in the narrower or the wider sense. Second, and this applies to turnover research and practice the like, we showed that there are different patterns of turnover, also depending on the sector. Thus, considering where employees potentially come from and are likely to go to will provide more insights and allow, for example, for more targeted measures to prevent loss of skilled employees. Further, it would be interesting to assess the effects of employee mobility and intellectual property (IP) turnover on the organizational performance from a network perspective.

Limitations

As an empirical study our research is not without limitations. First of all, the data collection approach might have resulted in a biased data set. We cannot be sure whether we have captured all essential companies in the Information Economy. However, we started the collection process with a very large industry-specific seed, and our 'snowball' system converged quickly on a core set of companies, which is a good sign. Second, the employees represented with personal profiles on the online professional network that we used as our data source might not be representative of the whole population in the Information Economy. This argument might be true for 'brick and mortar' industries, but in our context we should be able to assume that the average employee is Internet-savvy and having an online profile is a matter of course. Still, not all employees of a company will be present online. This is why we are mostly referring to relative comparisons between industries and companies, but do not rely on absolute numbers. The third and last limitation is imposed because our online professional network of choice did only provide turnover statistics for the top 5 inbound and outbound ties per company. As a result, smaller turnover numbers might not be present in our data. However, as we did start our data collection process with a large sample of companies and then followed the inbound and outbound ties, we do have a good mix of companies. This triangulation resulted in many more ties than the top 5 only as the following IBM example illustrates: in our network the company has 1600 outbound and 1098 inbound ties. Overall, we do not claim to have a complete picture, but we are very confident that our data is sufficiently valid and reliable to support our findings.

Conclusion

In this research effort, we studied turnover and employee mobility in the Information Economy. Our results provide new and unique insights into the ICT, Content and Media, and Internet sectors. We show that these sectors have very distinct characteristics in terms of employee mobility, as well as company size, type and age. This is why we conclude that the Information Economy is not a homogenous industry, at least not from a turnover perspective, which has some implications for research and practice. We understand this study as a first step in further providing a novel method to classify sectors and sub-sectors of the Information Economy based on real-world data, more precisely employee turnover data. Future research can build on these findings and continue to investigate in at least two areas: First, continue to explore differences between the sectors and sub-sectors in terms of actual turnover behavior. Second, review existing turnover literature and refine those findings taking into consideration the industry context where necessary.

References

Adomavicius, Gediminas, Bockstedt, Jesse C., Gupta, Alok, and Kauffman, Robert J. "Technology roles and paths of influence in an ecosystem model of technology evolution." *Information Technology and Management* 8, no. 2 (2007): 185–202.

Agrawal, A., Cockburn, I., and McHale, J. "Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships." *Journal of Economic Geography* 6, no. 5 (2006): 571–591.

Arlanis, Antonin, and Ciriani, Stéphane. "How Firms Interact and Perform in the ICT Ecosystem?" *Communications & Strategies* 122, no. 79 (2010): 121–141.

Bloomberg. "Public Company Search." 2013. http://investing.businessweek.com/research/common/symbollookup/symbollookup.asp, accessed December 2013.

Chiasson, Mike W., and Davidson, Elizabeth. "Taking Industry Seriously in Information Systems Research." *MIS Quarterly* 29, no. 4 (2005): 591–605.

Collet, Francois, and Hedström, Peter. "Old Friends and New Acquaintances: Tie Formation Mechanisms in an Interorganizational Network Generated by Employee Mobility." *Social Networks (*2013).

Crowston, Kevin, and Myers, Michael D. "Information Technology and the Transformation of Industries: Three Research Perspectives." *The Journal of Strategic Information Systems* 13, no. 1 (2004): 5–28.

CrunchBase. "CrunchBase - Home." 2013. http://www.crunchbase.com/, accessed December 2013.

Eriksson, R., and Lindgren, U. "Localized mobility clusters: impacts of labor market externalities on firm performance." *Journal of Economic Geography* 9, no. 1 (2008): 33–53.

Fortunato, Santo. "Community Detection in Graphs." *Physics Reports* 486, 3-5 (2010): 75–174.

Fransman, Martin. The New ICT Ecosystem: Implications for Policy and Regulation.

Cambridge, UK, New York: Cambridge University Press, 2010.

Garcia-Murillo, Martha, and MacInnes, Ian. "The Impact of Technological Convergence on the Regulation of ICT Industries." *International Journal on Media Management* 5, no. 1 (2003): 57–67.

Hair, Joseph F., Black, William C., Babin, Barry J., and Anderson, Rolph E. *Multivariate Data Analysis - A Global Perspective*. 7th ed. Upper Saddle River, NJ: Pearson, 2010.

Hallikas, Jukka, Varis, Jari, Sissonen, Heli, and Virolainen, Veli-Matti. "The evolution of the network structure in the ICT sector." *International Journal of Production Economics* 115, no. 2 (2008): 296–304.

Henrich, Joseph, Heine, Steven J., and Norenzayan, Ara. "Most People are not WEIRD."

Nature 466, no. 7302 (2010a): 29.

Henrich, Joseph, Heine, Steven J., and Norenzayan, Ara. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33, 2-3 (2010b): 61–83.

International Standard Industrial Classification of all Economic Activities (ISIC), Rev. 4.

New York: United Nations, 2008.

Karmarkar, Uday S., and Apte, Uday M. "Operations Management in the Information Economy: Information Products, Processes, and Chains." *Journal of Operations Management* 25, no. 2 (2007): 438–453.

King, William R., and Jun He. "External Validity in IS Survey Research." *Communications of the Association for Information Systems* 16 (2005): 880–894.

Koski, Heli, and Sierimo, Carolina. "Entry and exit in the ICT sector: New markets, new industrial dynamics?" *ETLA Discussion Papers* No. 847, 2003.

Lancichinetti, Andrea, and Fortunato, Santo. "Community detection algorithms: A comparative analysis." *Physical Review E* 80, no. 5 (2009).

Lee, Allen S., and Baskerville, Richard L. "Generalizing Generalizability in Information Systems Research." *Information Systems Research* 14, no. 3 (2003): 221–243.

Lee, Allen S., and Baskerville, Richard L. "Conceptualizing Generalizability: New Contributions and a Reply." *MIS Quarterly* 36, no. 3 (2012): 749–A7.

Martin, R., and Sunley, P. "Deconstructing clusters: chaotic concept or policy panacea?"

Journal of Economic Geography 3, no. 1 (2003): 5–35.

OECD Guide to Measuring the Information Society 2011: OECD Publishing, 2011.

Ritala, Paavo, Hallikas, Jukka, and Sissonen, Heli. "Coopetitive networks in the ICT sector."

International Journal of Business Environment 2, no. 1 (2008): 1.

Rosvall, Martin, and Bergstrom, Carl T. "Maps of Random Walks on Complex Networks reveal Community Structure." *PNAS* 105, no. 4 (2008): 1118–1123.

Rosvall, Martin, and Bergstrom, Carl T. "Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems." *PLoS ONE* 6, no. 4 (2011).

Seddon, Peter B., and Scheepers, Rens. "Towards the Improved Treatment of Generalization of Knowledge Claims in IS Research: Drawing General Conclusions from Samples." *European Journal of Information Systems* 21, no. 1 (2012): 6–21.

Tsang, Eric W. K., and Williams, John N. "Generalization and Induction: Misconceptions, Clarifications, and a Classification of Induction." *MIS Quarterly* 36, no. 3 (2012): 729–748.

Wasserman, Stanley, and Faust, Katherine. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.

Weick, Karl E. "What Theory is Not, Theorizing Is." *Administrative Science Quarterly* 40, no. 3 (1995): 385–390.

Xue, Ling, Ray, Gautam, and Gu, Bin. "Environmental Uncertainty and IT Infrastructure Governance: A Curvilinear Relationship." *Information Systems Research* 22, no. 2 (2011): 389–399.

Appendix: Tables and Figures

Employee Mobility as a Percent of Grand Total (Percentage Deviation of Observed Tie Distribution from Uniform Distribution)

From/To	ICT	Content/Media	Internet	Total
ICT	86.2% (+19.91%)	1.0% (-8.49%)	2.8% (-4.18%)	89.9% (+7.24%)
Content/ Media	0.7% (-8.87%)	6.2% (+7.66%)	0.3% (-0.73%)	7.1% (-1.94%)
Internet	1.3% (-6.12%)	0.3% (0.93%)	1.4% (+1.75%)	2.9% (-5.30%)
Total	88.2% (+4.92%)	7.4% (-1.76%)	4.4% (-3.16%)	100.0% (0%)

Table 1: Employee Mobility Across and Within the Information Economy Sectors

Information Economy **ICT Sector** **Content and Media** **Internet Sector**

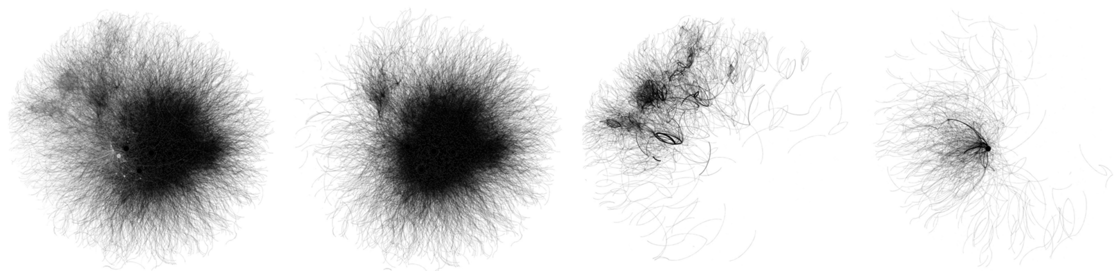
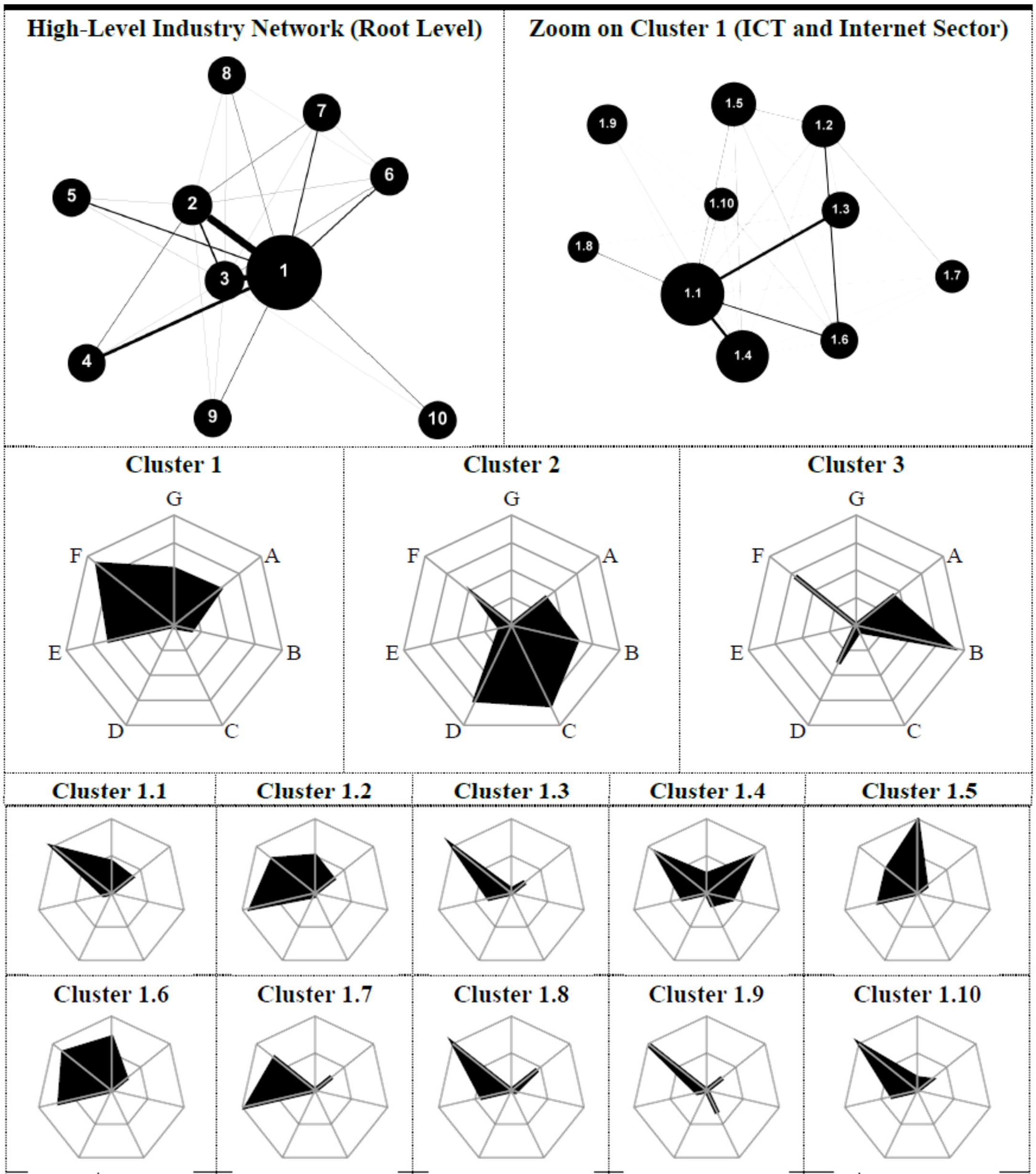


Figure 1: The Information Economy and Its Sectors



Letter	ISIC Description	ISIC Code
A	Internet	N/A
B	Publishing activities	J-58
C	Motion picture, video and television, sound recording and music	J-59
D	Programming and broadcasting activities	J-60
E	Telecommunications	J-61
F	Computer programming, consultancy and related activities	J-62
G	Manufacture of computer, electronic and optical products	C-26

Table 2: Clustering Results

Clu ster	Description	Examples	Comp anies	Run. Total	Int. Flow
High-Level Industry Clusters (Root Level)					
1	ICT and Internet	IBM, Oracle, HP, Accenture, Ericsson, Cognizant, Microsoft, Intel, Electronic Arts	11580	84.02%	99.03%
2	Entertainment Media	NBC Universal, Viacom, NBC, AMC, MTV, Fox, Vimeo, HBO, ABC, Disney	749	89.45%	90.20%
3	Press and Publishing	Thomson Reuters, Bloomberg, The Wall Street Journal, Businessweek, Pearson	547	93.42%	82.06%
4	Electronics	Schneider Electric, GE, ABB, APC, Areva, Industrial Electric, Wesco, Osram	81	94.01%	73.30%
5	Local Media	TVE, Antena 3, Telemadrid, El-Mundo, EFE, Premiere AG, Sky Italia	90	94.66%	77.80%
6	Online Media	CBS Interactive, CNET, ZDNet, Gamespot	116	95.50%	70.71%
7	Live Events	Live Nation, Ticketmaster Entertainment, AEG Live, House of Blues, Ticketfly	87	96.13%	54.72%
8	Local Media	Danmark Radio, TV2, Strix Television, Metronome, Nridsk Film, Egmont	73	96.66%	86.16%
9	Local Media	Hindustan Times, India Today, NDTV, Indian Express, Midday	46	97.00%	75.44%
10	Software Solutions	Systems Limited, Salsoft Technologies, Folio 3	58	97.42%	71.26%
Zoom on Cluster 1 (ICT and Internet Sector)					
1.1	Software and Platforms	IBM, Oracle, HP, Accenture, EMC, SAP, Dell	1001	8.64%	58.15 %
1.2	Mobile Equipment	Ericsson, Nokia, Alcatel-Lucent	414	12.22 %	60.44 %
1.3	Consulting, Outsourcing	Cognizant Technology Solutions, Tata Consulting Services, Infosys	244	14.33 %	58.65 %
1.4	Internet	Microsoft, Google, Apple, Yahoo, Amazon, Adobe, Ebay, Expedia, Facebook, Twitter	688	20.27 %	44.81 %
1.5	Semiconducto	Intel, AMD, Nvidia, Marvell, CSR, ARM	458	24.22	65.73
1.6	Network Equipment	Cisco, Juniper, Brocade, Aruba Networks	248	26.37 %	28.91 %
1.7	Wireless Provider	Verizon, T-Mobile, Virgin Mobile	132	27.51 %	65.25 %
1.8	Consulting	Cap Gemini, Logica, Atos, Steria	79	28.19	29.92
1.9	Games	Electronic Arts, Ubisoft, Codemasters, Activision	340	31.13	82.30
1.10	IT-Security	Symantec, McAfee, Trend-Micro, VeriSign	143	32.36	32.74

Table 3: Cluster Details

Paper V

Bibliographic data	De Oliviera, J. M., Zylka, M. P., & Gloor, P. A. & Joshi, T. (2019). *Mirror, Mirror on the Wall, Who is Leaving of Them All - Predictions for Employee Turnover with Gated Recurrent Neural Networks.* In: Studies on Entrepreneurship, Structural Change and Industrial Dynamics. Springer, Cham, https://doi.org/10.1007/978-3-030-17238-1_2
Status of publication	Published as completed research paper.
Publication type	Book chapter
Version & Copyright notice	Accepted version
Previous versions	Zylka M. P., de Oliveira J. M., Gloor P. A., Joshi T. & Tickoo, P. (2018). *Mirror, Mirror on the Wall, Who Is Leaving of Them All.* In: International Conference on Collaborative Innovation Networks Proceedings, Suzhou, China

Mirror, Mirror on the Wall, Who is Leaving of Them All - Predictions for Employee Turnover with Gated Recurrent Neural Networks

Employee turnover is a serious issue for organizations and disrupts the organizational behavior in several ways. Hence, predicting employee turnover might help organizations to react to these mostly negative events with e.g. improved employee retention strategies. Current studies use a "standard analysis approach" (Steel 2002) to predict employee turnover, accuracy in predicting turnover by this approach is only low to moderate. To address this shortcoming, we conduct a deep learning experiment to predict employee turnover. Based on a unique dataset containing twelve months of time series of e-mail communication from 3952 managers our model reached an accuracy of 80,0%, a precision of 74.5%, a recall of 84.4%, and a Matthews correlation coefficient value of 61.5%. This paper contributes to turnover literature by providing a novel analytical perspective on key elements of turnover models.

Introduction

Employees or so-called human resources are as important as other resources involved in the production of goods and services. They have a major effect on the productivity of a firm and are responsible for the creation of new knowledge – which may have a significant impact on firm growth. It is not surprising, therefore, that organizations seek to gain benefits by investing in human resources. These benefits, though, diminish when highly skilled employees voluntarily leave an organization for, say, a better position or higher wages at another organization. Further, while losing a highly skilled employee to a competitor may lead to access to external knowledge for the former employer, it also increases the risk that the former employer's knowledge will be leaked to the competitor (Somaya et al. 2008; Aime et al. 2010). Further, employee turnover reduces human capital of the former employer (Shaw et al. 2005) and is generally found to negatively impact company performance (Hancock et al. 2013). Hence, the competition for highly skilled personnel and their retention is critical to organizations. Such competition was dubbed the "war for talent" (Chambers et al. 1998) back in 1998 and later extended to the "war for Internet talent" (Efrati and Tam 2010), considering developments after the dot-com bubble burst in 2001.

Consequently, the turnover of highly skilled employees is a topic of interest for practitioners and scholars alike. It is a well-studied phenomenon in applied psychology, human resource, and general management literature (for an overview of turnover research history, see Holtom et al. 2008). Especially, the causes of actual turnover behavior are focused by research communities, since March and Simon (1958) proposed a first turnover model that takes the reasons for participating in organizations into account. They suggest that employees who are happy with their job and who do not have a job alternative will not leave their employer. Over thousands of scientific articles about employee turnover later, job dissatisfaction became and still is an important proximal cause for employee turnover (Lee and Mitchell 1994). Job dissatisfaction leads to thoughts of quitting and culminates in actual turnover behavior. Hence, it is important to know why and when highly skilled employees become dissatisfied with their current job.

To predict such changes and thus turnover behavior, researchers follow a "standard research design" (Steel 2002): Data on turnover predictors are collected via a survey and the information on the actual leaving is collected later. Then, the analysis is conducted with ordinary least squares (OLS) regression, survival and hazard functions or structural equation modeling (SEM). The accuracy in predicting turnover by this approach is only low to moderate (Lee et al. 2017), slightly higher using turnover intention as a proxy for actual turnover behavior. Novel approaches are necessary that take the trajectories of turnover determinants over time into account because changes over a specific period time in distal precursors affect turnover through changes in proximal antecedents (Hom et al. 2017; Lee et al. 2017).

In this paper, we follow the call for novel approaches for employee turnover prediction applying a gated recurrent neural network model (Chung et al. 2013), a variation of long short-term memory (LSTM) networks (Hochreiter and Schmidhuber 1997; Gers 2001) on longitudinal turnover data. Our goal is to predict turnover of highly skilled employees. Our research questions are:

- Is a deep learning model suitable for employee turnover prediction?
- How does the model perform in the prediction of employee turnover?

The remainder of this paper is structured as follows. The next section provides the theoretical grounding. This section contains an overview of related works regarding employee turnover, the role of job (dis-)satisfaction on employee turnover and the importance of the relational perspective of employee turnover. The subsequent section describes this study's setting, data, and methodology. The paper concludes with a discussion of the results and implications for future research.

Related Work

This section begins with a brief overview of employee turnover literature, the role of job satisfaction and communication behavior of employees on turnover. We then provide an overview of the idea behind recurrent neural networks (RNNs), long short-term memory (LSTM) and gated recurrent units (GRUs), all deep learning architectures that are important for understanding our study's experimental settings.

Employee Turnover

The investigation of employee turnover has a long history. March and Simon were the first who define a formal theory of turnover focused on participation in organizations (March and Simon 1958; Lee et al. 2017). Further research focuses on antecedents of individual employee turnover (Mobley 1982). Other research considers external factors such as promotional opportunities and kinship responsibility influencing employee turnover (Price and Mueller 1981). Moreover, studies focus on the negative and positive effects of employee turnover (Dalton and Todor 1979; Mobley 1982). In addition to the obvious negative effect on company performance, turnover can have some other negative effects. These can be differentiated into direct costs (e.g., for recruiting and training the successor) and indirect costs. The latter include loss of firm-specific human capital, demoralization of remaining employees, and loss of social capital embedded in the employees' relationships (Ton and Huckman 2008). Not surprisingly, loss of social capital from employee turnover negatively impacts company performance (Shaw et al. 2005). But, the opposite can also be the case: employee turnover can potentially lead to the creation of a business tie between the former and new employer, resulting in increased social capital for both firms (Somaya et al. 2008; Corredoira and Rosenkopf 2010).

More recent research on turnover analyzes the reasons for staying with a company and not the reasons for leaving (Lee et al., 2017). This shift in research objectives is based on the idea of job embeddedness, which incorporates off-the-job and on-the-job factors that embed employees in their job positions (Mitchell et al. 2001). Links to other colleagues, fit within the organization, and potential sacrifices in case of a turnover were found to influence intentions to leave an organization (Mitchell et al. 2001). Especially, relationships with other colleagues and groups are important for high job embeddedness and thus negatively related to turnover intention (Maertz and Griffeth 2004). The analysis of such relations between employees is mostly conducted by a social network analysis approach using centrality measures (c.f. Feeley 2000; Mossholder et al. 2005; Oldroyd and Morris 2012). Brass (1981) found no relationship between being central to an organization's workflow network and job satisfaction. However, more recent studies show that strong and positive social intra-organizational networks reduce employee turnover (Mossholder et al. 2005; Moynihan and Pandey 2007; Hom and Xiao 2011). Mossholder et al. (2005) found that network centrality might even predict employee turnover. Gloor et al. (2017) analyze the relational view of employee turnover in a detailed manner. Therefore, they analyze e-mail communication of 866 managers with the largest set of network and text analysis metrics so far and show that managers who quit have lower closeness centrality, less engaged conversations, shift their communication behavior starting from 5 months before leaving by increasing their degree and closeness centrality, the complexity of their language, as well as their oscillations in betweenness centrality and the number of nudges they need to send to peers before getting an answer. Gloor et al. (2007) show that intraorganizational e-mail communication analyzed with social network and text analysis metrics might be a promising predictor for turnover from the relational perspective.

However, the predictive power of past studies of employee turnover that follow the “standard research design” (Steel 2002) is low (Lee et al. 2017). Hom et al. (2017) and Lee et al. (2017) call for research that considers the dynamic nature of antecedents of employee turnover and conduct additional network-based investigations.

In the past years, several studies tried different methodologies to predict voluntary employee turnover and to overcome the issue of low predictive power (see Table 1 for review results). Especially, machine learning-based classification algorithms like Random Forests (RF), Gradient Boosting Machines (GBM), Support Vector Machines (SVM), Naïve Bayes (NB), k-Nearest Neighbors (kNN) are applied to employee turnover data typically gathered from human resource information systems (HRIS) or questionnaire surveys. Predictors are mostly socio-demographic and work-related items like working experience, tenure, skills, and performance. These studies present good accuracy values and claim that the machine learning-based classification algorithms outperform logit models (e.g. Nagadevara et al., 2008; Punnoose & Ajit, 2016; Zhao et al., 2019).

Our study follows the call of Lee et al. (2017) and is influenced by the study Gloor et al. (2017). However, we want to go one step further in turnover prediction by using recurrent neural networks (RNNs) that might enhance the predictive power of the relational perspective of employee turnover.

Research	**Study context**	**Turnover context**	**Data source**	**Sample size**	**Methodology**	**Predictor**	**Best model performance**
Chang (2009)	Employees in manufactory industries	Voluntary turnover	Not specified	881	Taguchi Method + Nearest Neighbor Classification Rules	Demographic factors, skillset, working experience	Accuracy = 0.878%
Hong et al. (2005)	Marketing specialists of Motor marketing company, Taiwan	Voluntary turnover	Not specified	132	SVM, Logistic Regression, Probit Model	Constant job performance	Accuracy = 0.84 with SVM
Nagadevara et al. (2008)	Software company, India	Voluntary turnover	Secondary data provided by the company	150	Neural Network, Logistic Regression,	Demographic factors, skillset,	Accuracy = 0.85 with C5.0

					Classification and Regression Trees, Classification Trees (C5.0), Discriminant Analysis	working experience	
Punnoose & Ajit (2016)	Leadership team of a global retailer, USA	Voluntary turnover	Provided by HRIS	73,115	Logit Model, NB, RF, kNN, LDA, SVM, XGBoost	Demographic factors, skillset, working experience	AUC = 0.86 with XGBoost
Quinn et al. (2002)	Workers of a regional child welfare agency, USA	Turnover behavior	Provided by internal IS	536	Logit Model, Multi-Layered Perceptron	Demographic factors, skillset, working experience	Accuracy = 0.60 (test set) with Multi-Layered Perceptron
Ribes et al. (2017)	Managers	Voluntary turnover	Not reported	1,000	NB, LDA, SVM, RF	Demographic factors, skillset, working experience	ROC = 0.95 with Random Forest
Sexton et al. (2005)	Workers of a small mid-west manufacturing company USA	Voluntary turnover	Reviewing personnel files	447	Neural Network Simultaneous Optimization Algorithm	Demographic factors, skillset, working experience, job level	Accuracy 0.96 with Neural Network
Sikaroudi et al. (2015)	Workers of an automotive parts manufacturing company, Iran	Voluntary turnover	Provided by HRIS	Not reported	Multilayer perceptron, RF, Probabilistic neural network, SVM, Classification and regression tree, kNN, NB, Apriori, CN2 algorithm	Demographic factors, skillset, working experience	Accuracy = 0.906 with Random Forest
Somers (1999)	Hospital employees, USA	Voluntary turnover	Questionnaire survey	577	Multilayer perceptron, Learning Vector Quantization	Affective commitment, Continuance commitment, Job satisfaction, Job	Accuracy = 0.88 with Multilayer perceptron

						withdrawal intentions	
Suceendran et al. (2015)	leading IT organization	Attrition	exit interview details of employees	2,572	ID3, J48, NB, Bayesian Network, K Star, IBK, Random tree, RF	Demographic factors, skillset, working experience, performance	Accuracy = 0.97 with IBK and Random Tree
Tzeng et al. (2004)	Nurses, Taiwan	Turnover intention	Questionnaire survey	389	SVM	Working motivation, job satisfaction, and stress levels	Accuracy = 0.892 with SVM
Zhao et al. (2019)	Employee of a regional bank, USA	Turnover behavior	Dataset 1: not reported Dataset 2: IBM Watson Database	Dataset 1: 9,089 Dataset 2: 1,470	Decision tree, RF, GBM, XGBoost, Logit model, SVM, neural networks, LDA, NB, kNN	Demographic factors, skillset, working experience, performance	ROC = 0.9008 with Extreme Gradient Boosting

Table 1. An overview of the literature on the application of machine learning algorithms in the context of employee turnover

Recurrent Neural Networks (RNN) & Gated Recurrent Units (GRU)

Recurrent neural networks (RNNs) are artificial neural networks that encode knowledge dependencies learned over past events and use this knowledge to reason about current events. RNNs achieve this by using cycles in a network, which allows past knowledge to persist in the form of inputs to the next network. RNNs can be thought of as a directed graph where each node passes the knowledge on to the next node after applying set of weights and transformations. Hence, RNNs are appropriate for data in the form of sequence or data with temporal attributes.

In recent years, RNNs have been used successfully in a variety of contexts, like object detection (Szegedy et al. 2013), speech recognition (Graves et al. 2013) or classification problems (Krizhevsky et al. 2012). Standard RNNs have a few limitations. One of them is that they cannot process inputs with varying length because long-term knowledge of too long input sequences cannot be stored (vanishing gradient problem). To process long-range sequences and identify relevant reasons for employee turnover, we need a model that can remember long-term knowledge. Long short-term memory (LSTM) is an RNN architecture that avoids the vanishing gradient problem (Hochreiter and Schmidhuber 1997; Gers 2001) and learns tasks that require knowledge of events that happened lots of time series earlier (Schmidhuber 2015). Therefore, we use a LSTM architecture with a simpler gating mechanism called gated recurrent units (GRU). GRUs are comparable to LSTM in terms of performance and exhibit better performance on smaller datasets (Chung et al. 2014).

Experiments

In this section, we present a brief description of our data, describe the data pre-processing steps, continue with the experimental settings, and conclude this section by presenting the results of our experiments.

Data Pre-Processing

This study is based on employees' e-mail communication data that is provided by a global professional services firm with more than 70,000 employees in over 20 countries. The e-mail data contain 845,208 actors (employees[1] and external stakeholders), ~138 million edges representing the communication that took place from 1st January 2017 till 29th January 2018.

We filtered the 845,208 actors by employees in managerial positions, which led to e-mail data of 3,952 managers containing 35% (48 million of 138 million edges) of the provided e-mail communication data. After collecting the mailboxes of the 3,952 managers, we prepared the data for the deep learning experiments. Therefore, we split the data in time frames (15 days time window with three days separation for the next window) and calculated network metrics (see Table 2) for each time frame with Condor[2], a social network and text analysis software. Each time frame was exported in a separate CSV file. The calculated network metrics are capable to show the changes in

---

[1] Some employees had more than one e-mail account in this company. In these cases, we merged the multiple e-mail accounts of an employee into one.

[2] http://guardian.galaxyadvisors.com

communication patterns of employees (c.f. Gloor et al. 2017) and hence serve as input variables for the deep learning models.

Metric	**Definition**
Messages sent	# of e-mails sent by an employee.
Messages received	# of e-mails received by an employee.
Messages total	Sum of e-mails sent and received by an employee.
Words total	# unique words an employee used in her e-mails.
Degree Centrality	# of colleagues each employee is directly connected within the communication network (Freeman 1979).
Betweenness Centrality	Likelihood to be on the shortest path between any two actors in the network. Indicates the extent to which each employee acts as an information hub and controls the information flow (Freeman 1979).
Betweenness Centrality Oscillation	# of local maxima and minima in the betweenness curve of an actor. Indicates how frequently employees change their network position in the team, from central to peripheral, and back.
Closeness Centrality	Inverse of distance of an actor from all other actors in the network, considering the shortest paths that connect each pair of actors. Indicates the efficiency of transmitting information and independence from other peripheral actors (Friedkin 1991).
Reach 2	# of colleagues each employee can reach by the distance of two.
Contribution Index	# messages sent - # messages received / (# messages sent + # messages received). Indicates how balanced a communication is in terms of sent and received messages.
Contribution Index Oscillation	# of local maxima and minima in the contribution index curve of an actor.
Ego ART	Avg # of hours sender takes to respond to emails. Time until a frame is closed for the receiver after she has sent an e-mail. Indicates the respect the receiver has for the sender.
Ego Nudges	Avg # of follow-ups that the sender needs to send to receive a response from the receiver.
Alter ART	Avg # of hours receiver takes to respond to emails. Time until a frame is closed for the sender, after she has sent an email. Indicates the respect the receiver has for the sender.
Alter Nudges	Avg # of follow-ups that the receiver needs to send to receive a response from the sender.
Avg. influence per message	Avg # of terms per message that has been introduced into the network.

Metric	Definition
Total influence	# new terms which a sender has introduced into the network and which are subsequently used by other members of the network. Indicates the extent to which someone causes the other person's pattern of speaking to match their own pattern.
Avg. Sentiment	Uses automatically generated bag of word, based on a dictionary trained for language/subject area. Indicates positivity and negativity of communication.
Avg. Emotionality	Standard deviation of sentiment. It represents the deviation from neutral sentiment.
Avg. Complexity	Information distribution using TF/IDF, independent of single words. Indicates the complexity of word usage. The more diverse words, which are all used evenly, a sender uses, the higher the complexity.

Table 2. Calculated network metrics that are considered as input variables for the machine learning models

For the machine learning experiments, we consider each CSV file as a single data point, meaning we combine 30 CSV files to have network metrics related to 3 months of e-mail communication as the input. Depending on the date an employee left the company, we consider her as a leaver or not. The last three months of work of employees who left the company are not considered in the model, because employees in this company are asked to send the resignation letter three months before leaving. Following Gloor et al. (2017), we assume that the closer employees get to the final decision of quitting, the higher the likelihood to exhibit divergent communication behaviors. Hence, we define five to seven months prior turnover as the period, where an employee is thinking about leaving the company. This means that 150 days before the actual turnover job satisfaction might turn to job dissatisfaction (see Figure 1).

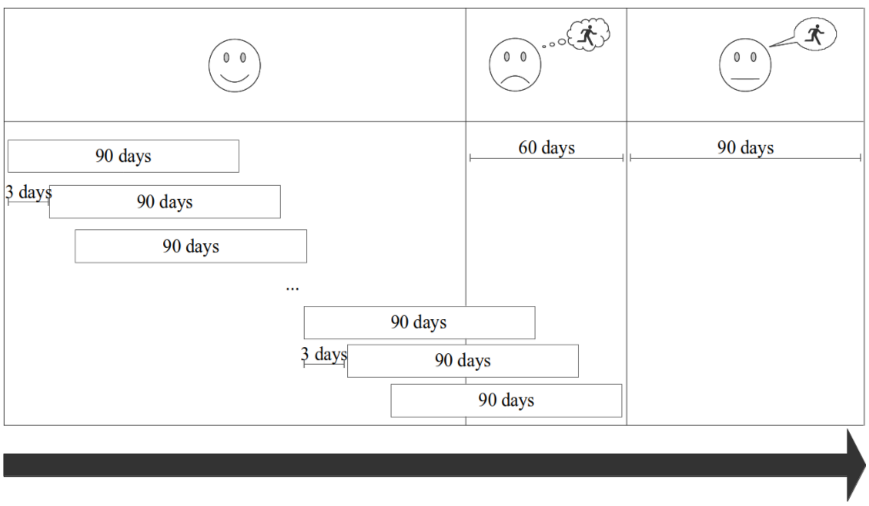


Fig. 1. Employee satisfaction timeline

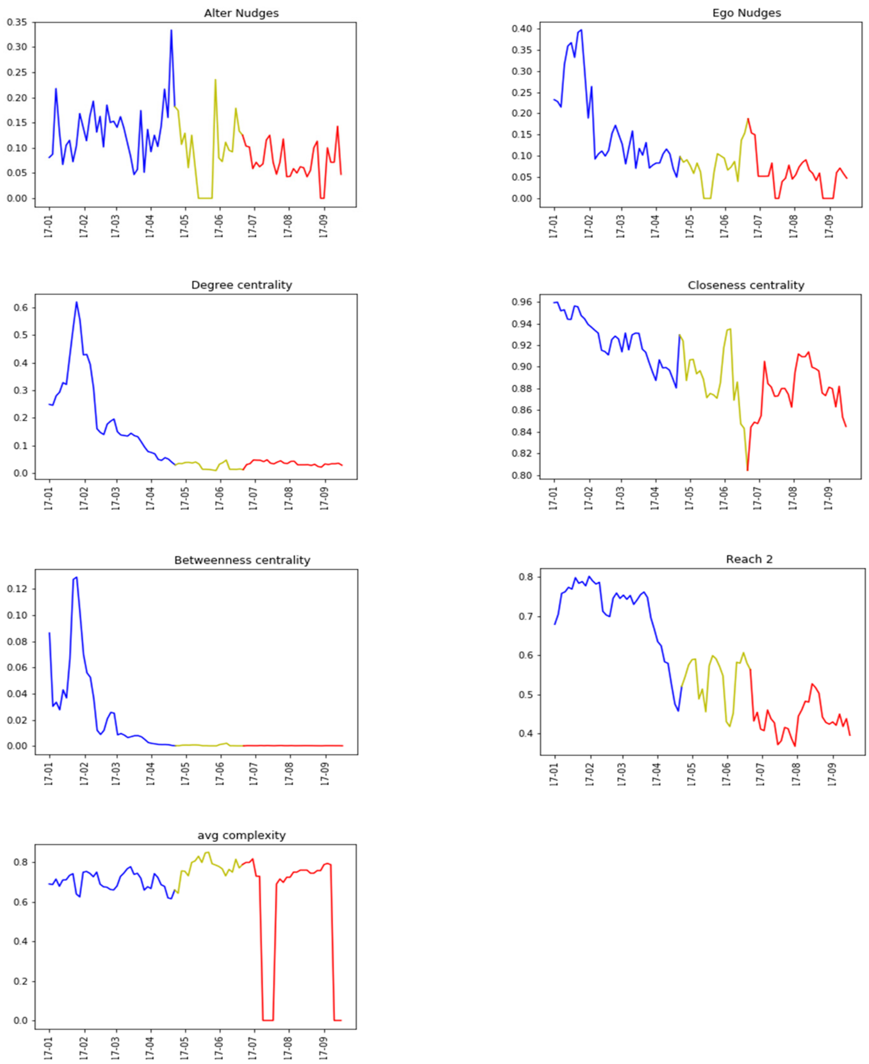


Fig. 2. An exemplary illustration of selected metrics of an employee. The x-axis indicates the time (monthly scale) and the y-axis indicates the values of the metric. The three phases of the employee satisfaction timeline are color-coded in these charts. Blue indicates the period in which the employee is satisfied. Yellow shows the phase in which the employee thinks about leaving the company. Red indicates the phase in which the employee has decided to leave the company and has signed the resignation letter

Experiment Setup

We built an RNN with GRU. We trained our model on 2898 employees and tested on the remaining 954 employees. Further, we apply 4-fold cross-validation to provide a robust estimate of the performance of our model. Therefore, we split the training data into four equal sized subsets, each subset has a similar number of employees who left (leavers). One of the four subsets are retained as the validation data for testing and the remaining three subsets are used for training. Since our dataset contains only a small number of actual employees who left the company (78 leavers), we had to preprocess the data with Synthetic Minority Over-Sampling Technique (SMOTE) to handle the imbalanced dataset (see Table 3 for dataset statistics). SMOTE is an over-sampling approach in which the minority class, in our case the leavers, is over-sampled. But instead of over-sampling with replacements, synthetic instances are created by joining any of the k minority class nearest neighbors of a minority class. These synthetic instances lead to more sensitivity regarding a minority class without too severe minority over-sampling.

We evaluate the results using precision, recall, accuracy, area-under-curve (AUC) and Matthews correlation coefficient (MCC) score, a discretization of the Pearson correlation value. We chose MMC as an additional performance measure because our model represents a binary classification problem (leavers / stayers) and MCC is more informative than other confusion matrix measures (such as F1 score and accuracy) in evaluating binary classification problems. MCC takes the balance ratios of the four confusion matrix categories (true positives, true negatives, false positives, false negatives) into account (Chicco 2017).

	Training	**Test**
# of actors	2898	954
# of leavers	59	19
# of stayers' time series	238438	78468
# of leavers' time series	832	262
# of stayers' SMOTE time series	238438	78468
# of leavers' SMOTE time series	238438	78468

Table 3. Statistics of the dataset

Results

Table 4 (training set) and Table 5 (test set) show four fold cross validation results of our dataset. We conducted the GRU experiment with a keep probability of 0.1, one LSTM layer, eight neurons and a learning rate of 0.001. All folds perform well, the average fold has an ACC of 0.933 and an MCC of 0.873. The test set's performance is good. The ACC is 0.800 and MCC (0.615) shows a strong positive predictive power.

Fold #	**ACC**	**P**	**R**	**AUC**	**MCC**
0	0,949	0,997	0,911	0,954	0,901
1	0,960	0,995	0,930	0,962	0,921
2	0,911	0,999	0,850	0,924	0,835
3	0,911	1,000	0,853	0,926	0,835
1.5 (avg.)	0.933	0.998	0.886	0.942	0.873

ACC=Accuracy, P=Precision, R=Recall, AUC=Area under curve, MCC=Matthews correlation coefficient

Table 4. Model results, training set

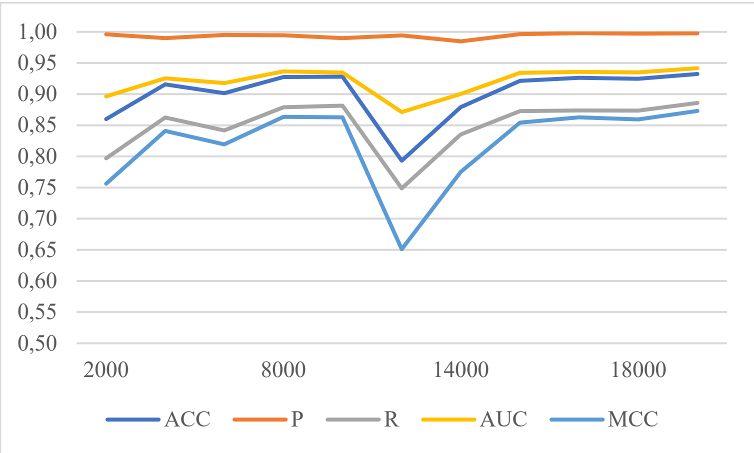


Fig. 3. Development of model performance indicators by model steps, training set, avg. fold

Fold #	ACC	P	R	AUC	MCC
0	0,796	0,700	0,867	0,808	0,604
1	0,740	0,547	0,891	0,782	0,520
2	0,883	0,962	0,830	0,893	0,775
3	0,780	0,770	0,786	0,780	0,561
1.5 (avg.)	0,800	0,745	0,844	0,816	0,615

ACC=Accuracy, P=Precision, R=Recall, AUC=Area under curve, MCC=Matthews correlation coefficient

Table 5. Model results, test set

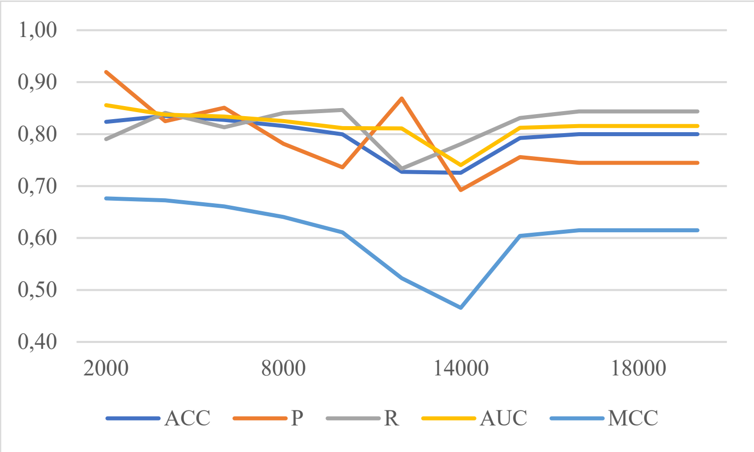


Fig. 4. Development of model performance indicators by model steps, test set, avg. fold

Table 6 shows the performance of the two GRUs with different configurations. Model 2 with sixteen neurons performs slightly better (MCC = 0.615) than Model 1 with eight neurons (MCC = 0.554).

	Configuration parameters	ACC	P	R	AUC	MCC
Model 1	Keep Probability: 0.1 Layers: 1 Size: 8 Learning Rate: 0.001 Type: 'GRU'	0.768	0.653	0.849	0.786	0.554
Model 2	Keep Probability: 0.1 Layers: 1 Size: 16 Learning Rate: 0.001 Type: 'GRU'	0.800	0.745	0.844	0.816	0.615

Table 6. Model results based on different configuration (test set, 1.5 fold)

Discussion and Implications

The results of the experiments revealed that applying a deep learning approach has potential to conduct a binary classification of employees in stayers or leavers by analyzing their e-mail communication behavior.

What are the implications from this study's findings for future RNN experiments and research on employee turnover? First, our study provides a novel approach to analyze longitudinal employee turnover data. This study is among the first to apply an RNN besides the usual applications of RNNs, such as object or speech recognition. As such, this paper may provide an insightful account for researchers interested in the context of employee turnover and provide an example of how deep learning methodology can be meaningful integrated in management research studies in general. We found it encouraging that the experiments went well in regard of performance. Further, we address the call by Hom et al. (2017) as well as Lee et al. (2017) by considering the dynamic nature of antecedents of employee turnover and conducting a network-based analysis in comparison to earlier research that primarily used a standard research approach (Steel 2002).

For organizations, the findings suggest the critical importance of human resource (HR) data analytics. We provide a possibility to predict employee turnover. HR managers can use our experiment setup as an 'early warning system' for employee turnover. Since employee turnover might be dysfunctional and has serious impact on company performance, HR managers could counteract with retention strategies when an important and highly skilled employee intend to leave the company. Nevertheless, ethical issues should be considered before applying this methodology in an organization.

Limitations and Future Research

One limitation of this work pertains to the generalizability of our proposed approach. It is plausible that the insights from this study might not directly apply to other companies or occupational groups. However, that would be a practical concern caused by insufficient data, which should be manageable.

Second, the premise of this study is that dissatisfaction derived from e-mail communication behavior culminates in leaving. Other options or paths of dissatisfied employees are ignored. Employees may lower job inputs or improve their circumstances (via promotion) rather than leave (with or without job offers) (Hulin et al. 1985). Thus, leaving is only one option among many ways to cope with job dissatisfaction.

Third, employee turnover data is an imbalanced dataset by nature. The number of leavers is always much lower than the number of stayers. Our dataset is extremely imbalanced, but we overcome this issue with SMOTE. However, a higher sample of leavers might improve the predictive performance of our models.

Fourth, this paper takes the relational perspective on employee turnover. However, traditional antecedents like organizational commitment were not considered.

Future research should conduct a classifier performance comparison by including several configurations of LSTMs, GRUs and other classifier models like Support Vector Machines (SVMs). Additionally, a comparison with models that are based on the mentioned standard research approach is necessary. Further, this study should be replicated with employee turnover data from other organizations, other occupational groups and with additional input variables.

Conclusion

In this paper, we applied a GRU RNN classifier that classifies employees in two states (leaver or stayer) by taking their e-mail communication behavior into account. The classifier's performance is measured in terms of confusion matrix with accuracy, recall, precision, AUC and MCC values. The developed GRU RNN model provides promising performance. Here, GRU can strongly benefit from the fact that it can look back in time and learn to correlate the calculated network metrics. GRU can learn these correlations, although it might require further training or different variables to be added to the data. In the future, we will continue to improve the performance of our model and conduct in depth error analysis.

We finally conclude that GRU is very suitable for classifying employees' turnover behavior. This is the first reported demonstration of a successful application of GRU neural networks to data from an organizational management context, namely employee turnover.

References

Aime F, Johnson S, Ridge JW, Hill AD (2010) The Routine May Be Stable But the Advantage is Not: Competitive Implications of Key Employee Mobility. Strateg Manag J 31:75–87. doi: 10.1002/smj.809

Brass DJ (1981) Structural Relationships, Job Characteristics, and Worker Satisfaction and Performance. Adm Sci Q 26:331. doi: 10.2307/2392511

Chambers EG, Foulon M, Handfield-Jones H, et al (1998) War for Talent. McKinsey Q 44–57

Chang HY (2009) Employee Turnover: A Novel Prediction Solution with Effective Feature Selection. WSEAS Trans Inf Sci Appl 6:417–426

Chicco D (2017) Ten Quick Tips for Machine Learning in Computational Biology. BioData Min 10:35. doi: 10.1186/s13040-017-0155-3

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

Corredoira RA, Rosenkopf L (2010) Should Auld Acquaintance Be Forgot? The Reverse Transfer of Knowledge Through Mobility Ties. Strateg Manag J 31:159–181. doi: 10.1002/smj.803

Dalton DR, Todor WD (1979) Turnover Turned Over: An Expanded and Positive Perspective. Acad Manag Rev 4:225–235.

Efrati A, Tam P-W (2010) Google Battles to Keep Talent. In: Wall Str. J. http://www.wsj.com/articles/SB10001424052748704804504575606871487743724. Accessed 5 May 2014

Feeley TH (2000) Testing a Communication Network Model of Employee Turnover Based on Centrality. J Appl Commun Res 28:262–277. doi: 10.1080/00909880009365574

Gers F (2001) Long Short-Term Memory in Recurrent Neural Networks. Lausanne, EPFL 2366:102. doi: 10.1162/neco.1997.9.8.1735

Gloor PA, Fronzetti Colladon A, Grippa F, Giacomelli G (2017) Forecasting Managerial Turnover Through E-Mail Based Social Network Analysis. Comput Human Behav 71:343–352. doi: 10.1016/j.chb.2017.02.017

Graves A, Mohamed A, Hinton G (2013) Speech Recognition with Deep Recurrent Neural Networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp 6645–6649

Hancock JI, Allen DG, Hancock E, Pierce CA (2013) Meta Analytic Review of Employee Turnover As a Predictor of Firm Performance. J Manage 1–56

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1–32. doi: 10.1162/neco.1997.9.8.1735

Holtom BC, Mitchell TR, Lee TW, Eberly MB (2008) 5 Turnover and Retention Research: A Glance at the Past, a Closer Review of the Present, and a Venture into the Future. Acad Manag Ann 2:231–274. doi: 10.1080/19416520802211552

Hom PW, Lee TW, Shaw JD, Hausknecht JP (2017) One Hundred Years of Employee Turnover Theory and Research. J Appl Psychol

Hom PW, Xiao Z (2011) Embedding Social Networks: How Guanxi Ties Reinforce Chinese Employees' Retention. Organ Behav Hum Decis Process 116:188–202. doi: 10.1016/j.obhdp.2011.06.001

Hong WC, Pai PF, Huang YY, Yang SL (2005) Application of Support Vector Machines in Predicting Employee Turnover Based on Job Performance. Adv Nat Comput, LNCS 3610, 668-674

Hulin CL, Roznowski M, Hachiya D (1985) Alternative Opportunities and Withdrawal Decisions: Empirical and Theoretical Discrepancies and an Integration. Psychol Bull 97:233–250. doi: 10.1037/0033-2909.97.2.233

Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst 25 1097–1105

Lee TW, Hom PW, Eberly MB, Mitchell TR (2017) On the Next Decade of Research in Voluntary Employee

Turnover. Acad Manag Perspect 31:201–221. doi: 10.5465/amp.2016.0123
Lee TW, Mitchell TR (1994) An Alternative Approach: The Unfolding Model of Voluntary Employee Turnover. Acad Manag Rev 19:51–89. doi: 10.5465/AMR.1994.9410122008
Maertz CP, Griffeth RW (2004) Eight Motivational Forces and Voluntary Turnover: A Theoretical Synthesis with Implications for Research. J Manage 30:667–683. doi: 10.1016/j.jm.2004.04.001
March JG, Simon HA (1958) Organizations. Wiley-Blackwell, Cambridge, MA
Mitchell TR, Holtom BC, Lee TW, et al (2001) Why People Stay: Using Job Embeddedness to Predict Voluntary Turnover. Acad Manag J 44:1102–1121. doi: 10.2307/3069391
Mobley WH (1982) Some Unanswered Questions in Turnover and Withdrawal Research. Acad Manag Rev 7:111–116
Mossholder KW, Settoon RP, Henagan SC (2005) A Relational Perspective on Turnover: Examining Structural, Attitudinal, and Behavioral Predictors. Acad Manag J 48:607–618. doi: 10.5465/amj.2005.17843941
Moynihan DP, Pandey SK (2007) The Ties that Bind: Social Networks, Person-Organization Value Fit, and Turnover Intention. J Public Adm Res Theory 18:205–227. doi: 10.1093/jopart/mum013
Oldroyd JB, Morris SS (2012) Catching Falling Stars: A Human Resource Response to Social Capital's Detrimental Effect of Information Overload on Star Employees. Acad Manag Rev 37:396–418. doi: 10.5465/amr.2010.0403
Punnoose R, Pankaj A (2016) Prediction of Employee Turnover in Organizations Using Machine Learning Algorithms. Int J Adv Res Art Intel 9:22-26
Price JL, Mueller CW (1981) A Causal Model of Turnover for Nurses. A Causal Model of Turnover for Nurses. Acad Manag J 24:543–565. doi: http://dx.doi.org/10.2307/255574
Quinn A, Rycraft JR, Schoech D (2002) Building a Model to Predict Caseworker and Supervisor Turnover Using a Neural Network and Logistic Regression. J Technol Hum Serv 19:65–85
Schmidhuber J (2015) Deep Learning in Neural Networks: An Overview. Neural Networks 61:85–117. doi: 10.1016/j.neunet.2014.09.003
Sexton RS, McMurtrey S, Michalopoulos JO, Smith AM (2005) Employee Turnover: A Neural Network Solution. Comput Oper Res 32:2635–2651
Shaw JD, Duffy MK, Johnson JL, Lockhart DE (2005) Turnover, Social Capital Losses, and Performance. Acad Manag J 48:594–606. doi: 10.5465/AMJ.2005.17843940
Sikaroudi AME, Ghousi R, Sikaroudi AE (2015) A Data Mining Approach to Employee Turnover Prediction (Case Study: Arak Automotive Parts Manufacturing). J Ind Syst Eng, 8:106-121.
Somaya D, Williamson IO, Lorinkova N (2008) Gone But Not Lost: The Different Performance Impacts of Employee Mobility Between Cooperators Versus Competitors. Acad Manag J 51:936–953. doi: 10.5465/AMJ.2008.34789660
Somers, MJ (1999) Application of Two Neural Network Paradigms to the Study of Voluntary
Employee Turnover. J Appl Psychol 84:177
Steel P (2002) At the Empirical Turnover Theory of Fit Interface: Problems. Acad Manag Rev 27:346–360. doi: 10.5465/amr.2002.7389900
Suceendran KM, Saravanan R, Ananthram D, Poonkuzhali S, Kumar RK, Sarukesi K (2015) Applying Classifier Algorithms to Organizational Memory to Build an Attrition Predictor Model.
Szegedy C, Toshev A, Erhan D (2013) Deep Neural Networks for Object Detection. Neural Inf Process Syst Conf 2553–2561
Ton Z, Huckman RS (2008) Managing the Impact of Employee Turnover on Performance: The Role of Process Conformance. Organ Sci 19:56–68. doi: 10.1287/orsc.1070.0294

Tzeng HM, Hsieh JG, Lin YL (2004) Predicting Nurses' Intention to Quit With a Support Vector Machine: A New Approach to Set up an Early Warning Mechanism in Human Resource Management. Comput Inf Nurs 22:232–242

Zhao Y, Hryniewicki MK, Cheng F, Fu B, Zhu X (2018) Employee Turnover Prediction with Machine Learning: A Reliable Approach. In Proceedings of SAI Intelligent Systems Conference 737-758. Springer, Cham.

Paper VI

Bibliographic data	Gloor, P. A., Zylka, M. P., Fronzetti Colladon, A., Makai, M. (2022). *'Entanglement' – a new dynamic metric to measure team flow.* In: Social Networks, Vol.70, July, pp. 100-111, https://doi.org/10.1016/j.socnet.2021.11.010.
Status of publication	Published as completed research paper.
Publication type	Journal article
Version & Copyright notice	Accepted version Licensed under CC BY-NC-ND
Previous versions	n/a

‘Entanglement’ – a new dynamic metric to measure team flow

We introduce “entanglement”, a novel metric to measure how synchronized communication between team members is. This measure calculates the Euclidean distance among team members’ social network metrics timeseries. We validate the metric with four case studies. The first case study uses entanglement of 11 medical innovation teams to predict team performance and learning behavior. The second case looks at the e-mail communication of 113 senior executives of an international services firm, predicting employee turnover through lack of entanglement of an employee. The third case analyzes the individual employee performance of 81 managers. The fourth case study predicts performance of 13 customer-dedicated teams at a big international company by comparing entanglement in the e-mail interactions with satisfaction of their customers measured through Net Promoter Score (NPS). While we can only speculate about what is causing the entanglement effect, we find that it is a new and versatile indicator for the analysis of employees’ communication, analyzing the hitherto underused temporal dimension of online social networks which could be used as a powerful predictor of employee and team performance, employee turnover, and customer satisfaction.

Keywords: Dynamic social network analysis; synchronization; communication patterns; entanglement; group cohesion; flow state

Introduction

Albert Einstein called quantum entanglement “spooky action at a distance” (Einstein et al., 1935), predicting that quantum mechanics should allow objects to influence each other’s action at great distance. It took other Nobel prize winning physicists’ decades after Einstein’s death to confirm his prediction. In this paper we propose a similar social entanglement effect between people. Note that we are not making any conclusive claim about the cause of this social entanglement effect, we just find that it seems to exist, and posit that there seem to be useful parallels between quantum entanglement and social entanglement that assist in the conceptualization of the latter.

“You share everything with your bestie. Even brain waves.“ (Angier, 2018). This is how the New York Times summarized the work of Parkinson, Kleinbaum and Wheatley (2018), who found that brain scans of close friends show similar patterns as they watch a series of short videos. Using these results, the researchers trained a computer algorithm to predict the strength of a social bond between two people based on the relative similarity or synchronization of their neural response patterns. Such neural synchronization patterns are also observed in various other studies in different contexts, e.g., to determine neural contingencies between musical performers and their audiences. Hou et al. (2020) assess the neural synchronization between violinist and audience and the relation to popularity of violin performance. Their findings suggest that neural synchronization between the audience and the performer might serve as an underlying mechanism for the positive reception of musical performance. Further, neural synchronization can be confirmed by analyzing verbal group communication (Liu et al., 2019). Individuals try to achieve neural and body synchronization in order to facilitate fluid interaction (Fairhurst et al., 2013; Yun et al., 2012). Experiments show that synchrony of fingertip movement and neural activity between two persons increases after cooperative interaction (Yun et al., 2012). Hence, engaging individuals in synchronized activities like walking, dancing etc. is an effective way of increasing subsequent cooperation between those individuals.

However, the studies mentioned above focus on neural or body synchronization and are not applied in typical work environments or contexts. But “being in sync” or “in flow” in work environments is a relevant research topic and should be considered by decision-makers to determine the impact of such behavior on employee performance. Being in sync with others can increase cooperation by strengthening social attachment among team members (Wiltermuth and Heath, 2009). Thus, it might also affect team productivity and team performance positively. Offline and online communication plays an important role to distinguish between teams that are in sync or "out of sync". Where offline communication like face-to-face meetings establish team synchronization easily (Maznevski and Chudoba, 2000), online communication such as e-mail and chat tools might diminish team synchronization (Hinds and Bailey, 2003). The asynchronous characteristic of online communication, for instance caused by time lags (Cramton, 2001), may hinder developing a shared team rhythm (Hinds et al., 2015; Hinds and Bailey, 2003).

However, there exist opportunities to analyze online communication data in near-real time for continuous monitoring of team learning and performance. Metrics based on communication flow from person to person or amount of communication are suitable for real-time processing. In addition, studies have shown that analyzing online communication data in organizational contexts (de Oliveira et al., 2019; Gloor et al., 2017b) could be used as a predictor for job-related constructs, such as employee turnover or employee performance. Speed of responding to an e-mail, for example, is a good predictor of individual and team performance (Gloor et al., 2020). It might be a proxy for the passion of the person who is responding to an e-mail (Gloor, 2017), or for other external reasons such as urgency, power differentials, etc.

Based on these behavioral and neuroscientific insights and findings on the relationship between interpersonal synchronization and communication, we hypothesize that being in sync can also be shown by analyzing patterns of team online communication gathered through a social network analysis (SNA) approach. Hence, our research questions are:

(1) Are time series of communication patterns from online communication valid indicators for analyzing the synchronization of a team and its flow?
(2) Is a measure for team flow capable to predict job-related outcomes such as job performance or employee turnover?

We answer these questions by introducing a metric called *entanglement*, which measures the synchronization of e-mail communication behaviors of team members and their flow state over time. This metric is grounded in SNA and identifies the similarity of timeseries of SNA metrics. We validate the metric by conducting four case studies, with different datasets from different organizations. Each case study is in a different context and variants of the entanglement measure are used as a predictor of different individual and group performance indicators.

The rest of the paper is organized as follows. In Section 2 we present the theoretical background of flow state and team synchronization. Subsequently, we illustrate the idea of our entanglement metrics, which want to capture how much people interact in the same rhythm or are “in sync”. We finalize this section with the metric’s formalization. In Section 3, we explain the data collection and applied methods. Then in Section 4, we introduce four case studies in which we demonstrate the predictive power of the proposed entanglement metrics. In the last section, we discuss results and advocate future research.

Theoretical background

Team synchronization and flow state

Synchronization is a fundamental element of life. Besides neuronal synchronization mentioned in the introduction, one finds studies that deal with the synchronization of human activities (Guastello and Peressini, 2017). Synchronization is often defined as the manifestation of unintended coordination. It is part of the natural behavior of a human being and takes place so invisibly that we usually do not notice it. It is triggered by audio-visual stimuli, haptic perception or simply by the presence of certain people. Synchronization can be analyzed as neuromuscular coordination, where there is a relatively exact or proportional tracking of body, hand and head movements, autonomic arousal, or electroencephalogram (EEG) readings between two or more people (Guastello and Peressini, 2017). For example, Néda et al. (2000) show that the audience of a concert synchronizes its applause after an asynchronous start and Fairhurst et al. (2013) and Yun et al. (2012) show that people synchronize their finger tapping to improve coordination. While these studies only look at synchronization as neuromuscular coordination and task coordination, there are research efforts currently underway to uncover connections between synchronization in cognition, task structures, and performance outcomes in teams (Gipson et al., 2016). Better work performance outcomes would also be expected when teams are similarly synchronized (Elkins et al., 2009; Stevens et al., 2013).

The hypothesis that team synchronization leads to better performance is further motivated by the theory of *flow state.* While the concept of synchronization in the above-mentioned studies applies a natural science perspective, human sciences like positive psychology consider synchronization as a part of *flow state* (Gloor et al., 2012) and expect flow state to cause better performance. A team is in flow state (Csikszentmihalyi, 1996) when members create a sense of shared confidence and empathy, which culminates in a collective mental state in which individual intentions harmonize and are in-sync with those members of the group. This condition is also referred to as achieving a "group mind", which is marked by a deep emotional resonance which enables e.g., jazz musicians to be completely coordinated throughout the improvisational flow. In other words, group flow manifests itself in physical and verbal activities, for instance people mirroring each other and quickly finishing each other's sentences using the same words and phrases, indicating a "parallel synchronization of thought" (Armstrong, 2008). The more the team members are in-sync, the more likely it is to observe group flow. Group flow can be analyzed applying "interaction analysis", which entails closely observing and categorizing the interactions, movements, and body language of group members. But, it cannot be limited to neurological studies of particular participants of the group's emotional conditions or subjective memories (Sawyer, 2003). Thus, group flow cannot be split down into specific tasks; rather, it is a process that arises from group dynamics and has the ability to improve job satisfaction, intrinsic motivation, vigor, performance or efficiency (Delarue et al., 2008; Sawyer, 2003; van den Hout et al., 2018). Hence, flow represents rather an oscillating dynamic state that combines continuous and sudden changes across time (Ceja and Navarro, 2012) than a static one.

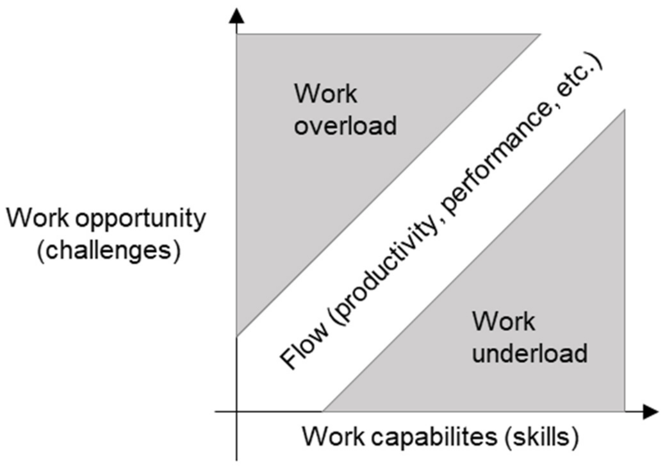


Figure 1. General model of flow state in work environments. Based on original model of flow state by Csikszentmihalyi (2000).

The flow concept can be transferred into the organizational context (Heyne et al., 2011). Bakker (2005) defines work-related flow as a short-term peak experience at work that is characterized by absorption, work enjoyment and interest. Teams "are in flow" if there is a certain balance between challenges and the skill sets of the individual team members. Work-related flow leads to a better productivity and performance (see Figure 1). Further, by the definition of flow by Csikszentmihalyi (1996) high flow leads to high performance. If a team is collectively in flow, it therefore will deliver high performance. In general, flow is likely to correlate positively with measurable results (Quinn, 2005). Quinn (2005, p. 611) emphasizes that "[i]n knowledge work [...] flow may be a useful concept for understanding performance.". Studies of flow proceed from a broader awareness that team processes like communication need to be studied as events over time (Arrow et al., 2004).

Entanglement conceptualization and formalization

The idea of the entanglement measure is to determine how a person is in sync with his/her group and shares the same flow with the other team members, with regards to communication over a period of time. In an attempt to conceptualize entanglement, a multidisciplinary approach is proposed, bringing together concepts from several disciplines, ranging from quantum mechanics to human and social sciences. The term entanglement is borrowed from quantum physics, where a pair or group of particles which are "entangled" mysteriously change their quantum state at the same time, even when the group of particles is physically far apart at different locations on the world (Horodecki et al., 2009). A result of this phenomenon is that when one measures the quantum state of one particle, one simultaneously determines the quantum state of the other particle. A quantum state (of a particle) is a representation of knowledge or information about an aspect of the system or reality (Pusey et al., 2012). In this study, we interpret the reality as the state about a person-to-person relationship. Thus, the two particles are seen as two individuals that have potentially interacted with "others", not necessarily with each other, and have therefore become entangled. Our idea of synchronicity is that people are in-sync when they show similar behavioral patterns, such as communication activity. Hence, two persons are entangled even when they are physically separated or not involved in a (local) interaction with each other but share a similar communication behavior (an example is provided in Figure 2).

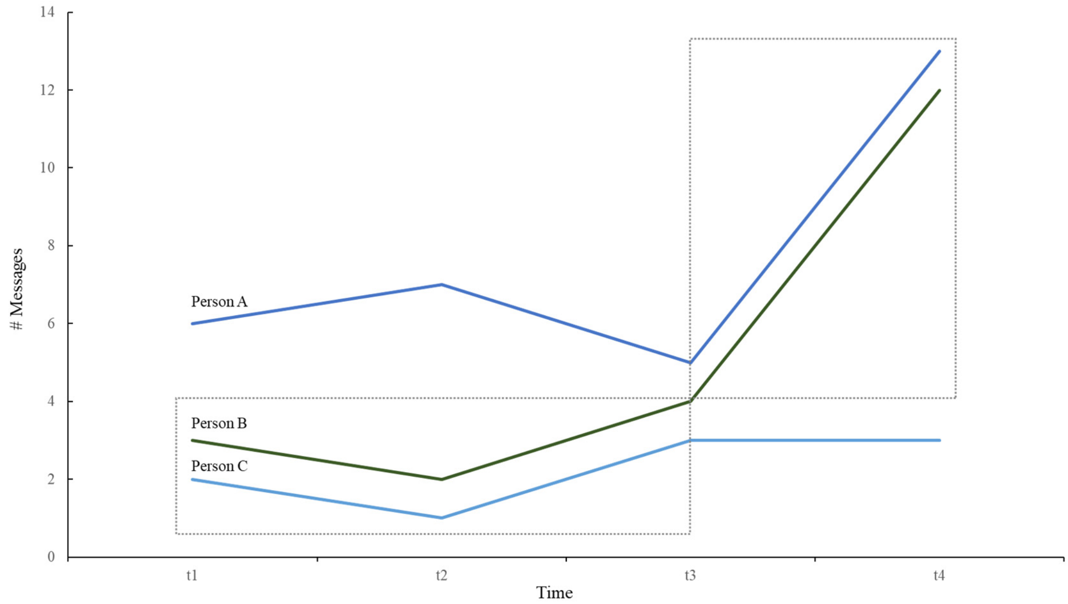


Figure 2. Communication intensity of three persons by time

Similar concepts have previously been described in psychology and sociology. “Entrainment” describes a process where one system’s motion or oscillation frequency synchronizes with another system, for instance the brainwaves of two people rocking together in their chairs. Cross et al. (2019) define interpersonal entrainment as the synchronization of organisms to a rhythm, for example singing, dancing, or even walking together. Much earlier, early twentieth century French sociologist Emile Durkheim defined collective effervescence as the similar but broader notion of synchronized action between humans (Durkheim, 2008), to describe when a community or society comes together to communicate the same thought or participate in the same action. This concept has been picked up by sociologist Randall Collins through his construct of “Interaction Ritual Chains” (Collins, 2005), which explain collective action through shared emotional energy. The common theme of all these constructs is colocation, people creating and experiencing emotional energy by being together at the same location. We therefore prefer the term “entanglement” to describe synchronous action between humans independent from where they are located, to describe in the words of Albert Einstein, “spooky action at a distance”.

Human communication is fundamentally synchronous and rhythmic, two important characteristics of individual and interactional behavior (Condon, 1986). The synchronization of interactional behaviors helps to generate a sense of flow state for the persons involved (Condon, 1986). Further, it always takes other people for a person to reach the state of flow (Collins, 2005), while the other people do not have to be physically present. Thus, entanglement leads to a flow state of two persons analogous to the "mysterious change" of a particle's quantum state. Intuitively, we propose that the "more similar the communication" of two persons A and B is, the more person A is in sync and is able to share the same flow of communication with person B over a period of time. Individuals that are in flow might have higher abilities to productively channel their cooperative spirit when working together.

Figure 2 shows the communication of three persons by time. Person B and C communicate in similar intensity (here: number of sent messages) from *t1* until *t3*. Their communication decreases from *t1* to *t2* and increases from *t2* to *t3* by the same amount. Further, their lines in the chart are very close together meaning the distance between each of their data points is short. We observe the same pattern for person A and person B in time period *t3* and *t4*. Such patterns might indicate synchronization. Thus, we can state that the distance of the data points representing the communication intensity between two or more persons in a specific time window is an indicator for their synchronization. Here, we use the Euclidean distance, a straight-line distance between two points in Euclidean space. We calculate the Euclidean distance d of two data points x and y of a communication metric A of the same time window t with:

$$d\big(A(x_t), A(y_t)\big) = \sqrt{\big(A(x_t) - A(y_t)\big)^2}$$

This Euclidean distance specified in the formula above is calculated for every pair of nodes and time window t. An essential requirement to determine if persons are entangled is to consider both team synchronization and team flow. Team flow is based on flow experienced in relational embeddedness (Burt, 2005) which can be established by e.g. communication and collaboration. To address this structural feature of communication, we propose to apply SNA. SNA offers a suitable methodology to study group dynamics as well as to investigate the role of the individuals within these dynamics (Wasserman and Faust, 1994). It focuses on various aspects of the relational structures and the flow of information, which characterize a network of people, through graphs and structural measures.

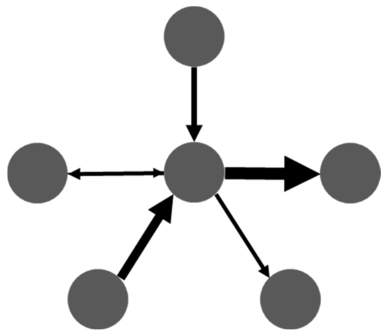


Figure 3. Graph representing an e-mail communication network

To better illustrate the concept of "entanglement" we consider an e-mail network, characterized as a graph made of a set of nodes (e-mail accounts) and a set of directed edges (weighted by the number of e-mails) connecting these nodes. The direction of an edge specifies the source (e-mail sender) and target (e-mail receiver) node; the weight of an edge shows the relation intensity (number of e-mails) between two nodes (see Figure 3). For example, if person A sends 3 emails to person B, we see an arc originating at node A and terminating at node B of weight equal to 3.

To illustrate the idea and calculation of entanglement with an example, we use an individual mailbox representing a dataset of e-mails of persons that work together on several projects. First, we collected the mailbox and stored it in a database, where the e-mail data was structured from a network perspective. In order to calculate the entanglement of the mailbox owner and his/her colleagues, we take the inverse of the Euclidean distance of the time series of the communication activity represented by messages sent over time for each node/actor in the network. This value will get the larger the more similar the activity time series of two actors are. However, we have to distinguish between two pairs of actors at different locations in the network, one pair embedded into a tight cluster communicating with many other actors, while the other pair is exchanging the same number of e-mails as the first pair, but is only weakly connected to other actors. To make this metric comparable among pairs of actors with different levels of activity in the same network, we multiply it by the product of the degree centralities of both actors. Degree measures the centrality, sometimes seen as a proxy of popularity, of a node in a network, by counting the number of its nearest neighbors (Freeman, 1978). Further it can be a proxy for the level of engagement within a group, team or organization (Gloor et al., 2020). Communication activity via e-mail (Gloor et al., 2014) indicates the number of e-mail messages sent by a person within a time interval.

Figure 4 shows the e-mail communication activity over a period of time, for the email box we analyzed. The blue line shows the mailbox owner's communication activity, the other lines correspond to the people s/he is exchanging e-mails most frequently with. The more correlated the communication activity between the owner of the mailbox and another person are, the more they are in sync, share the same flow over a period of time, and thus are entangled. The picture also illustrates the need to include degree centrality in the entanglement formula, as the levels of activities, while running in parallel, are vastly different for different people.

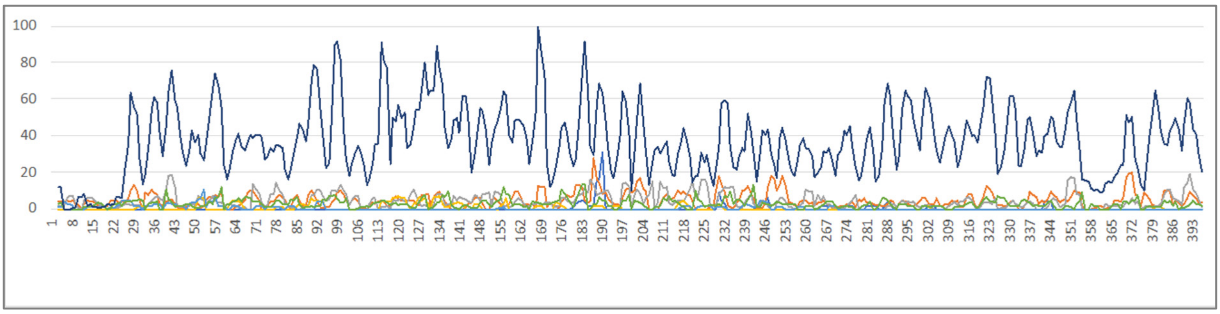


Figure 4. Flow of e-mail communication by time

Accordingly, we define the activity entanglement $E_A(x_T, y_T)$ between two individuals, named x and y in a specific time window T, as:

$$E_A(x_T, y_T) = \frac{C_D(x_T)\, C_D(y_T)}{d\big(A(x_T), A(y_T)\big)}$$

where $C_D(x_T)$ and $C_D(y_T)$ are the degree centralities of the two individuals x and y, and $d\big(A(x_T), A(y_T)\big)$ is their Euclidean distance, with respect to communication activity A in a defined time window T. In other words, the entanglement of two individuals x and y is given by the multiplication of the number of their direct contacts in the e-mail network divided by their synchronization of communication activity. As has been said above, it is necessary to include the product of the degree centralities of x and y into the entanglement formula to provide for the differences in centralities between actors: assume that actor x has low degree, if x is synchronized with highly connected actor y having high degree centrality, the high degree of actor y will boost entanglement of actor x in comparison with all other actors in the network. In other words, we want our metric to reward less influential actors that are synchronized with influential actors.

Similarly, we could consider not just communication activity, but also individuals' synchronization in weighted and unweighted betweenness centrality. Betweenness is a well-known metric in social network analysis. It is the sum of the fraction of all-pairs shortest paths that pass through a node v (Freeman, 1977):

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)},$$

where V is the set of nodes, $\sigma(s,t)$ is the number of shortest paths from s to t, and $\sigma(s,t|v)$ is the number of those paths passing through node v (Brandes, 2001). Inverse arc weights are considered for the determination of node distances. To control for network size, the above index is usually normalized between zero and one.

If the betweenness centrality time series of two individuals are in sync, it means that they share similar network positions, and levels of influence, at the same time. Individual betweenness entanglement E_B is the product of the degree of two individuals divided by their Euclidean distance in betweenness centrality over a period of time.

$$E_B(x_T, y_T) = \frac{C_D(x_T)\, C_D(y_T)}{d\big(C_B(x_T), C_B(y_T)\big)}$$

In addition, we speculate on the possibility to evaluate how much an individual is in sync with the aggregated flow of the entire network. As a proxy of the aggregated rhythm of the team we take Freeman's group betweenness centralization, C_{GB} (Freeman, 1978). Group betweenness centralization is the sum of the differences between the betweenness centrality of the most central node, $C_B(v^*)$, and that of all other nodes in the network (Freeman, 1978; Wasserman and Faust, 1994), normalized by its maximum value which is $(G-1)^2(G-2)$ where G is the total number of nodes:

$$C_{GB} = \frac{2\sum_{i=1}^{G}[\,C_B(v^*) - C_B(v_i)]}{[(G-1)^2(G-2)]}.$$

This definition of group betweenness centralization is appropriate for this use case, as we compare how entangled an individual node is with all other nodes with regards to betweenness.

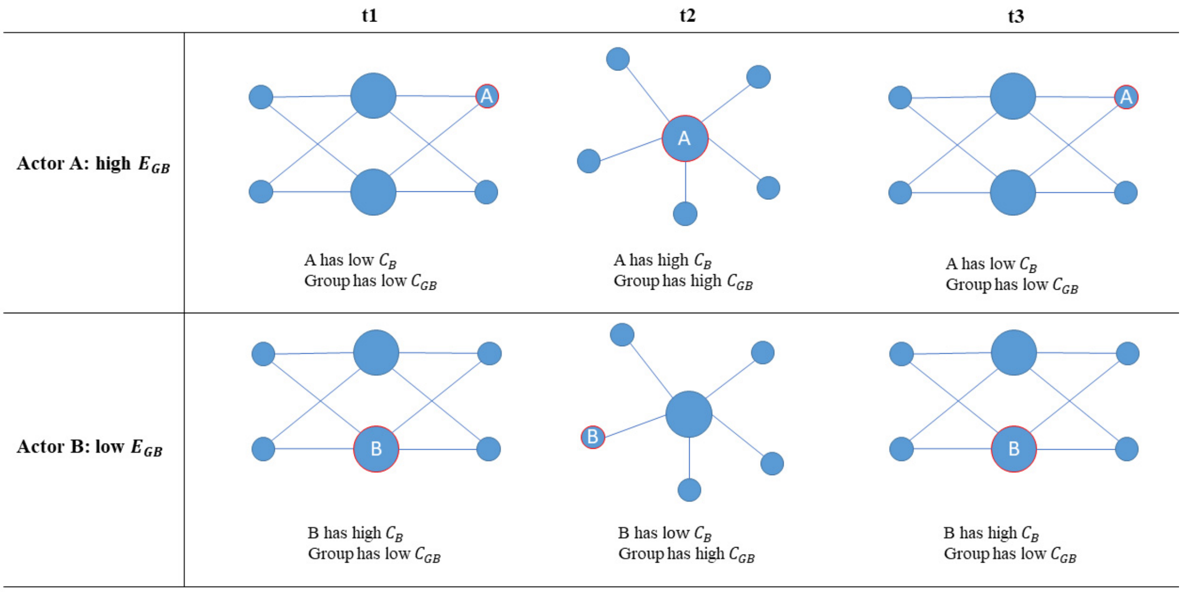


Figure 5. Intuitive motivation for group betweenness entanglement ($\boldsymbol{E_{GB}}$)

Figure 5 gives an intuitive motivation for the usefulness of group betweenness entanglement. It shows a group of six actors at three points in time of a changing network structure. Actor A is very much "entangled" with the overall group: In t1 and t3, when the group betweenness centralization (C_{GB}) is low, his/her (individual) betweenness centrality (C_B) is low also, in t2, when the group betweenness centralization is high, his/her C_B is high too, leading to low Euclidean distance of his/her C_B to C_{GB}, resulting in high entanglement. In contrast, actor B is lowly "entangled" with the group, in t1 and t3 when C_{GB} is low, his/her betweenness centrality (C_B) is high, in t2 when C_{GB} is high, his C_B is low. This leads to high Euclidean distance to C_{GB}, and thus to low entanglement. Formally, we measure group betweenness entanglement $E_{GB}$by dividing group betweenness centralization C_{GB} by the Euclidean distance of group betweenness centralization and betweenness centrality of the actor being analyzed over a time period. C_{GB_T} – as a metric of variation – is an indicator for the centralization of the group in time window T, the individual betweenness centrality $C_B(x_T)$ in this sense is an influence on C_{GB_T}, i.e., how much an actor impacts C_{GB_T}. Intuitively this metric reflects the contribution of this actor to the level of centralization of its group. In other words, it measures how far away the normalized betweenness centrality of an actor is from the betweenness centralization of its group at any point in time. If an actor's betweenness is high and its group betweenness centralization is high, the actor is probably responsible for the centralized network structure – thus the Euclidean distance between group betweenness centralization and an actor's betweenness centrality is small, and therefore the actor's group betweenness entanglement high. On the other hand, if an actor's betweenness is low and its group betweenness centralization is high, it means somebody else is central and the actor is unimportant in betweenness centrality terms, thus less entangled with the group. We look at this across groups (frequently analyzing advice networks in work settings) and over time. Accordingly, we define group betweenness entanglement, $E_{GB}(x_T)$ of x as:

$$E_{GB}(x_T) = \frac{C_{GB_T}}{d\left(C_B(x_T), C_{GB_T}\right)}$$

To show the inequality in individual group betweenness entanglement we calculate the Gini coefficient for E_{GB}:

$$G(E_{GB}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| E_{GB}(x_i) - E_{GB}(x_j) \right|}{2n^2 \overline{E_{GB}}}$$

The same formula can also be used for activity entanglement to calculate $G(E_A)$. Intuitively, the Gini coefficient measures inequality in the distribution of entanglement among all actors in a network. This is based on the observation that for an actor x being resource-poor or resource-rich in a network – the resource being entanglement in this case – can be highly predictive for the behavior or performance of x. It therefore makes sense to put the entanglement of x in relationship to the entanglement of all other actors in the network through Gini entanglement.

Data collection and methods

In this section, we present the data collection process and the methods we applied to analyze the data for the case studies. For each case, we ran the same data collection process. We fetched the e-mails of a sample of project members who chose to participate in each pilot study. All worked at large organizations at the time we collected their communication data. We used Condor[1], a social network and semantic analysis software to collect and analyze the data. We normalized the e-mail data for time zones. In our calculations we set the time window to 7 days, as this has been shown to deliver the best results for this type of organizational e-mail data (Gloor 2017).

We measured the relationship of entanglement calculated from e-mail communication with individual and group outcome variables. Since we explore the properties of communication networks, we focused on the calculation of communication-based measures – such as messages sent and received – and of network centrality measures, as we explained in section 2.2. Further, we used the reach-2 metric, which is the number of nodes that a social actor can reach by going through each of its direct links in the graph (Gloor, 2017). Reach-2 has been used as a proxy for social capital, as it measures the number of connections of the people a person is connected to (de Oliveira and Gloor, 2018).

In addition, we relied on online communication metrics developed specifically for assessing interactivity in e-mail communication. In particular, we looked at the communication activity (Gloor et al., 2014), which indicates the number of e-mail messages sent by a person within a time interval, and at the number of nudges, which represents the average number of pings (emails) that a sender needs to send in order to receive a response from the receiver (Gloor et al., 2014). Here we differentiate between ego nudges (the number of pings before a recipient responds) and alter nudges (the number of pings before others respond). In addition, we measured the contribution index which is the balance between messages sent and messages received (Gloor, 2017). Lastly, we calculated the average response times (ART) to measure how much time it takes a person to reply to an e-mail (Gloor et al., 2014; Merten and Gloor, 2010). This metric is helpful to identify fast and slow communicators and recognize patterns of behavior looking at periods of slower response. We separate between Ego ART, the average number of hours a sender takes to respond to e-mails and Alter ART, the average number of hours recipients takes to respond to a sender.

[1] http://www.ickn.org/ckntools.html
Condor is available for free academic use and includes the entanglement metrics.

Case studies

We illustrate, in four case studies, how the proposed *entanglement* metric can be used with e-mail data to predict work-related outcome variables, such as team performance and employee turnover (see Table 1). The four case studies we present here are related to different business contexts and consider different dependent variables. In all cases we analyze email data, illustrating the suitability of the entanglement metric for online communication data. Our goal here is not to directly compare results across case studies, deriving general conclusions, or claiming causality. Rather we want to show the versatility of our entanglement metrics, which can be adapted to study business interaction dynamics in different scenarios.

Case study	Industry	Research object	Entanglement measure	Entanglement level	Outcome variable
A	Health care	53 employees in 11 healthcare innovation teams	Activity entanglement	Team	Team performance & learning behavior
B	Professional services	113 senior executives	Activity entanglement	Individual	Employee turnover
C	Professional services	81 managers	Betweenness entanglement	Individual	Employee performance
D	Professional services	82 mangers in 13 teams	Group betweenness entanglement	Team	Customer satisfaction

Table 1. Case studies overview

Case study A – learning behavior and performance

This case study was conducted as a pilot in a health care organization to determine if activity entanglement E_A between 53 team members of 11 medical innovation teams could predict performance and learning behaviors.

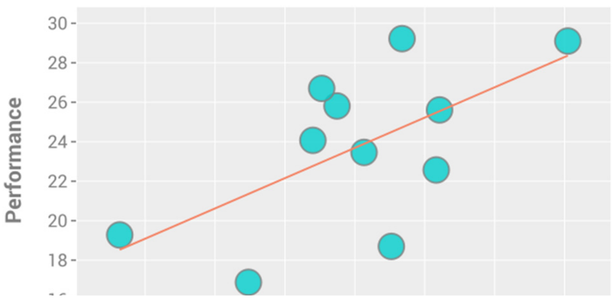


Figure 6. Entanglement correlation with performance

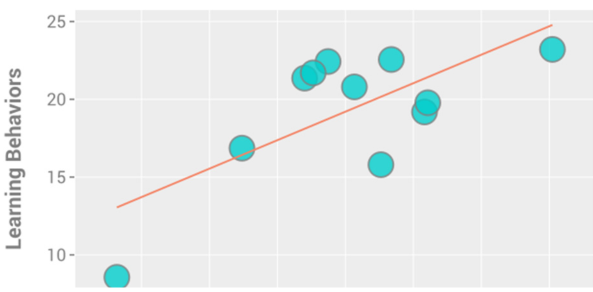


Figure 7. Entanglement correlation with learning behavior

The performance and learning behaviors of each team was rated and triangulated every other month for the duration of a year by three overall project managers. They individually rated the team performance and the capability of a team to learn new things. At the same time, all e-mails of the project members were collected and analyzed. Individual activity entanglement of each actor with all other actors was calculated, and then the average was taken for each actor. Finally, for each team average and standard deviation of activity entanglement over all team members was computed.

We find that team performance and learning behavior are significantly correlated with the standard deviation of activity entanglement of team members, as shown in Figure 6 and Figure 7 (which show a scatter plot of the two metrics, with a fitted regression line). The Pearson's correlation coefficient of the standard deviation of activity entanglement of team members with team performance is .615 ($p = .045$) and with learning behavior is .707 ($p = .015$). In other words, the wider the spread in activity entanglement E_A of the team members, the higher their performance and learning behavior. This pattern corresponds to a few core team members being strongly entangled, and the remaining members showing weak E_A. We also notice that moderate dispersion of entanglement is associated to higher variability in performance scores. This could be explained by control variables we could not collect in this study due to limited data availability. Alternatively, it could suggest that in order for performance to be high, few employees have to take a strong group lead, guiding the others towards a common goal.

Case study B – turnover prediction

In our second case study, we conducted a pilot study at a global professional services firm. In this case we wanted to evaluate the possible association of entanglement with executives' decision to leave the firm, through voluntary resignation. Turnover of highly important employees such as senior executives is critical for companies, because it has negative implications for firm performance (Hancock et al., 2013; Zylka and Fischbach, 2017).

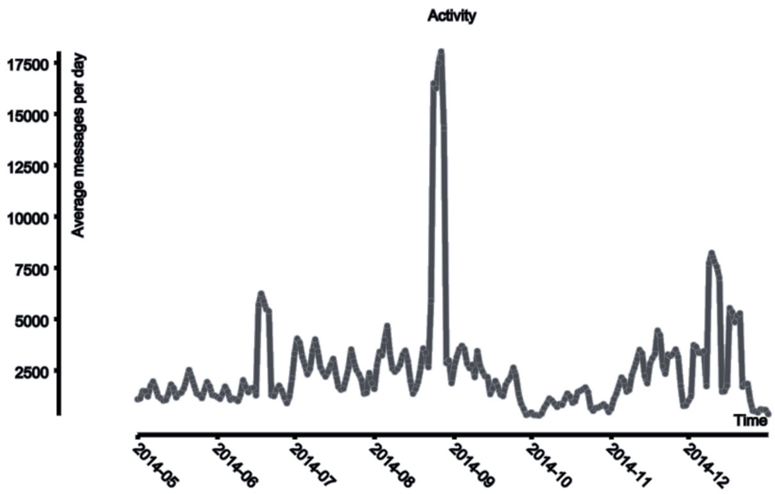


Figure 8. Communication activity by time

Eight months of e-mail data of 113 senior executives at a large global services company was collected from May to December 2014 (see Figure 8). We calculated activity entanglement E_A of 55 employees who left the firm from January to May 2015. To determine the inequality in entanglement, we also calculate the Gini index of E_A, for each person (from an ego perspective) in the network, considering her/his entanglement with all other peers. The Gini index measures the dispersion of entanglement scores of a social actor with all others in the network. In an "egalitarian" network with low Gini index for each node, all actors are either highly or weakly entangled, in a "non-egalitarian" network with high Gini index some actors are highly entangled, while others are weakly entangled. This was compared with the activity entanglement E_A of a control group made of 58 employees, who were selected randomly and still working in an unterminated position at the firm in June 2015.

From a preliminary t-test, we immediately notice that there is a significant difference in the Gini index of activity entanglement, between senior executives who leave the company (M = .457, SD =

.070) and those who stay (M = .488, SD = .059), t(111) = -2.513, p = .013. On average, Gini entanglement is significantly higher for those who stay.

Past studies have shown that managerial disengagement might depend on multiple factors and that communication-based and social network analysis metrics, captured from e-mail communication, can reveal it (Gloor et al., 2017b). Accordingly, we present Pearson's correlations (in Table 2) and logistic regression models (in Table 3), to see if the effect of the entanglement variable remained significant when combined with other predictors. The highest correlation of entanglement is with contribution index, which however does not lead to collinearity issues. A high contribution index is an indication for "spammers", the higher the contribution index, the more somebody sends compared to receiving e-mail. If there is a spammer, s/he will be entangled with many, while others who are sending much less, will thus be less entangled. This results in a high Gini entanglement for that person. Extending this effect to all users will lead to high correlation between the two values.

		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	**Leaver (1 = yes)**	1											
2	**Rank**	.056	1										
3	**Tenure**	.032	.067	1									
4	**TSLP**	-.018	-.012	.534**	1								
5	**Msg sent**	-.050	.114	.196	-.014	1							
6	**Msg received**	.040	.208*	.129	-.123	.632**	1						
7	**CI**	-.168	-.040	.131	.039	.626**	.219*	1					
8	**Reach 2**	.024	.306**	.208*	-.137	.431**	.554**	.310**	1				
9	**Betweenness**	.092	.221*	.185	.106	.445**	-.018	.279**	.236*	1			
10	**Alter ART**	.071	-.109	-.060	.231*	-.199*	-.216*	-.176	-.416**	-.060	1		
11	**Ego ART**	.233*	-.216*	.011	.066	-.227*	-.210*	-.388**	-.357**	-.066	.529**	1	
12	**Gini entanglement**	-.232*	.014	.083	-.044	.741**	.554**	.840**	.422**	.208*	-.243**	-.342**	1

* p < .05; ** p < .01.

Table 2. Correlations for leavers

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Rank	0.40740	0.30430	0.06247	0.32153	0.10214
Tenure	0.00146	0.00184	0.00094	-0.00039	-0.00403
TSLP	-0.00080	-0.00004	0.00012	0.00036	0.00148
Msg sent		0.00007	-0.00036	-0.00047	-0.00033
Msg received		0.00013	0.00049	0.00056	0.0016226**
CI		-0.91878	-0.76631	-0.23275	3.326418**
Reach 2			-0.00017	0.00074	0.00037
Betweenness			0.00004	0.00004	0.00005

Alter ART				-0.00313	-0.00891
Ego ART				0.021418*	0.0299733**
Gini entanglement					-35.02065**
Constant	-0.62053	-1.12106	-0.72954	-1.64071	16.28513**
Pseudo R-squared	0.00470	0.02930	0.04970	0.08420	0.17960

** p < .05; ** p < .01.*

Table 3. Logistic regression for leavers

We first tested a model with only the control variables of rank, tenure, and time since last promotion (TSLP) measured in months. In the subsequent models, we added the other predictors in blocks showing, in Model 4, that the only significant predictor, before adding entanglement, is Ego ART. This suggests that managers who leave the company are less responsive to e-mails and take more time to answer. In the full model, Ego ART, messages sent, contribution index and Gini activity entanglement are significant. Including this last predictor in the model leads to a significant improvement of the McFadden's pseudo-R-squared, which more than doubles (going from .08 to .18). As we can see from Model 5, a higher entanglement makes the probability of leaving the company smaller.

To evaluate the possibility of using the entanglement variable for making predictions, we used machine learning. In particular, we used a tree boosting model named CatBoost and its related Python library (Prokhorenkova et al., 2018). This boosting approach is now well-known and proved its usefulness in past research, where it also sometimes outperformed other supervised machine learning methods, such as Support Vector Machines (SVM) and Random Forest Models (Huang et al., 2019). The model performance has been assessed through Monte Carlo Cross Validation (Dubitzky et al., 2007), with 300 random splits of the dataset into train and test data (75% vs 25%). Thanks to the contribution of our variables, we could achieve an average accuracy of predictions of 80.25%, with an average value of the Area Under the ROC-Curve (AUC) of 0.81.

In a second step, we considered the average model resulting from cross-validation and used it to interpret the impact of each variable on predictions (calculated as the average of its absolute Shapley values). We used the *SHapley Additive exPlanations* (*SHAP*) Python package (Lundberg and Lee, 2017). This method proved to be particularly suitable for tree ensembles and to work well also with respect to other approaches (Lundberg et al., 2020, 2018). As Figure 9 shows, the Gini index of activity entanglement is the variable with the highest impact on model predictions. Its contribution is much higher than all other variables, again supporting the importance of this metric. At the second place, we find Ego ART. Results are consistent with those of logit models and indicate that managers who are slower in answering e-mails, and have low Gini entanglement, are more likely to leave the company. Low Gini entanglement means that they show constant levels of entanglement, either being entangled with almost nobody or everyone – a situation that might be stressful to maintain, especially when associated with email overload (Reinke and Chamorro-Premuzic, 2014). Average/high levels of Gini entanglement, on the other hand, have a positive impact on the prediction of staying in the company. This means that these managers show uneven entanglement, being highly entangled with some colleagues, while being weakly entangled with others. The probability of staying in the company is even higher for average entanglement values.

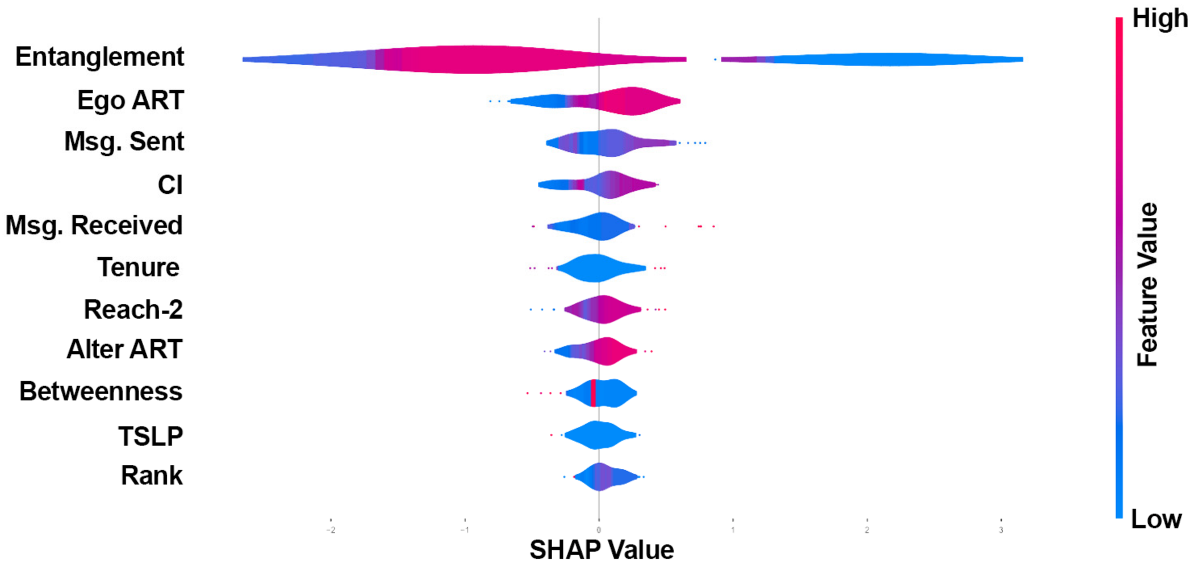


Figure 9. SHAP values (prediction of leavers)

Case study C – employee performance

We analyzed the e-mail interactions of 81 managers working for a big international services company. Every year the performance of managers was evaluated by their bosses and by the HR department. Whereas the rating of almost all of these managers was "exceeded expectations" for the year 2015, we noticed that 15 of them obtained a lower rating. Like in the case study B of resigning senior executives, we were interested in understanding if entanglement could be related to individual work performance. Carrying out a t-test, we could see that there is a significant difference between the Gini coefficients of betweenness entanglement E_B scores of top (M = .508, SD = .061) and low (M = .469, SD = .028) performers, $t(79) = 2.432$, $p = .017$.

As we did for leavers in case study B, we additionally built logistic regression models to assess the combined impact of variables on the probability to be a low performer. Pearson's correlations among our predictors are presented in Table 4. The highest correlation of entanglement is again with contribution index, but this time lower than case study B.

		1	2	3	4	5	6	7	8	9	10	11
1	**Low Performer (1 = yes)**	1										
2	**Tenure**	.238*	1									
3	**TSLP**	.153	.178	1								
4	**Msg Sent**	-.070	.175	.060	1							
5	**Msg Received**	.084	.080	.190	.186	1						
6	**CI**	-.074	.206	.066	.731**	.180	1					
7	**Reach 2**	-.038	.237*	.004	.200	.492**	.227*	1				
8	**Betweenness**	-.082	.157	-.039	.844**	-.031	.517**	.228*	1			
9	**Alter ART**	-.136	-.118	-.180	.154	-.171	.134	-.264*	.046	1		
10	**Ego ART**	-.139	-.155	-.188	-.024	-.125	.021	-.276*	-.068	.406**	1	
11	**Gini entanglement**	-.264*	-.085	.007	.484**	.208	.548**	.234*	.334**	-.024	.118	1

* p < .05; ** p < .01.

Table 4. Correlations for low performers

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Tenure	0.0105286*	0.011992**	0.0138258**	0.0134347**	0.0128212*
TSLP	0.01189	0.01115	0.01266	0.01019	0.00714
Msg Sent		-0.00030	-0.00166	-0.00170	-0.00196
Msg Received		0.00036	0.0017504*	0.0017746*	0.0020867*
CI		-0.44577	1.60707	1.69432	2.958886*
Reach 2			-0.00006	-0.00088	0.00144
Betweenness			-0.0104667**	-0.0101837**	-0.0085401*
Alter ART				-0.00636	-0.01227
Ego ART				-0.00733	0.00157
Gini entanglement					-26.67039*
Constant	-3.0535****	-3.608166***	-2.26197	-1.44970	10.88953
Pseudo R-squared	0.06980	0.09900	0.22400	0.23140	0.28030

** p < .10; ** p < .05; *** p < .01; **** p < .001.*

Table 5. Logistic regression for low performers

As Table 5 shows, in the full model the p-value of Gini entanglement is only < .1; however, the inclusion of this variable leads to a good improvement of the McFadden's pseudo-R-squared, from .2314 (Model 4) to .2803 (Model 5). A significant performance improvement is also obtained by including weighted betweenness centrality.

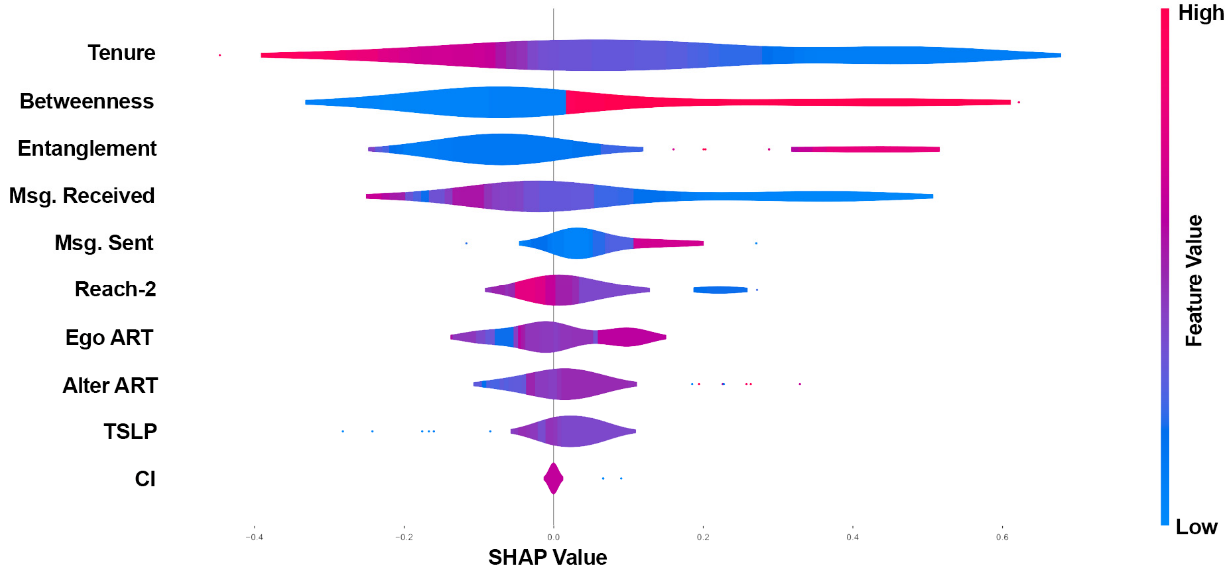


Figure 10. SHAP values (prediction of top performers)

The usefulness of the entanglement predictor is confirmed by the results of the CatBoost model that we trained to classify managers into top and low performers. We followed the same procedure as in the previous case study B – i.e., a Monte Carlo cross-validation with 300 repetitions – and obtained good average results (Accuracy = 74,73%, AUC = 0.68). Figure 10 shows the Shapley values associated with each predictor. For an easier reading, we coded top performers as 1 and low performers as 0 (here the model is predicting top performers, which is exactly symmetrical to the choice of predicting low performers that we did in Table 5). Tenure, betweenness centrality and entanglement are the most important predictors – with high Gini coefficient of betweenness entanglement and high betweenness centrality significantly increasing the chance of being classified as a top performer. These managers are highly entangled with some colleagues, and weakly entangled with others – demonstrating selective communication behavior with close collaborators, while being efficient with their time and communicating comparatively less with the rest of the organization.

Regarding tenure, we observe the opposite effect, with recently hired employees generally receiving better ratings.

Case study D – customer satisfaction

In this case study, we show that entanglement is significantly related to team performance, measured as customer satisfaction through the Net Promoter Score (NPS). 13 teams within the company participated to our study, comprising a total of 82 managers. Each team was dedicated to a specific client.

We measured betweenness entanglement of each team by taking the group betweenness entanglement of each member and considering group dispersion by means of the Gini coefficient.

We find that high group betweenness entanglement inequality is positively related to team performance – this time measured as customer satisfaction. Running a Pearson's correlation test, we find a significant association of Gini group betweenness entanglement with team performance ($r = .522$, $p = .002$). For each team we have repeated measures over three time periods. Therefore, we used multilevel linear models (Hoffman and Rovine, 2007; Nezlek, 2008; Singer and Willett, 2009) as a more appropriate technique to evaluate the possible effect of entanglement on customer satisfaction. We nested repeated measures into groups (level 2). Results are presented in Table 6.

	Model 1	**Model 2**
Gini Group Betweenness Entanglement		0.6418315*
Constant	0.5244776***	0.2869091*
Variance L2	0.0655537	0.0455178
Variance L1	0.020654	0.0211838
Variance Change L2		-30.56%
Variance Change L1		2.57%

*Note. * p < .05; *** p < .001.*

Table 6. Multilevel models for customer satisfaction (N = 34, with 13 groups)

As the table shows, the biggest variance proportion can be attributed to team characteristics: the intraclass correlation coefficient is 0.7604, meaning that 76% of the empty model variance is at level 2 (Model 1). Including the entanglement variable in the model (Model 2) reduces this variance of 30.56%, which is a highly significant result for a single predictor. The higher the inequality in group betweenness entanglement is, the happier the customer is. Similarly to case study A, this confirms that selective communication of teams, where some team members are highly entangled and others are not, leads to happier customers.

Discussion and conclusions

In this study, we propose a novel synchronization metric, called *entanglement*, which is based on SNA of e-mail communication between different actors. We demonstrate with four case studies on real-world datasets that this metric and its variants are a good predictor of different individual and team performance indicators (a summary of our results is provided in Table 7).

Case study	**Dependent Variable**	**Result summary**
A	Team performance and learning behavior	The wider the spread in activity entanglement of the team members, the higher the team performance and learning behavior. This corresponds to having some core team members strongly entangled and the remaining members weakly entangled.
B	Employee turnover	The Gini index of activity entanglement is the variable with the highest impact on model predictions. Employees who stay in the company have high Gini entanglement probably using selective communication and interacting more with some colleagues than with all others. They are also more responsive to emails and take less time to answer.
C	Individual performance	Tenure, betweenness centrality and Gini entanglement are the most important predictors of top performers. – with high Gini index of betweenness entanglement and high betweenness centrality significantly increasing the chance of being classified as a top performer.
D	Customer satisfaction	The Gini index of group betweenness entanglement for teams, is related to customer satisfaction. The higher the inequality in group betweenness entanglement is for a team, the happier its customer is. This means that customers are happier when a few entangled leaders emerge in the team.

Table 7. Case study results summary

Firstly, we find that dispersion of activity entanglement is positively associated with team performance. This means that the synchronized communication activity of some team members and their continuous similar flow state improve the performance of the team. These findings resemble studies showing that e-mail communication and face-to-face communication frequency (Patrashkova-Volzdoska et al., 2003), and flow in knowledge work (Quinn, 2005), can both lead to higher team performance. It also seems that the best teams exhibit higher dispersion, comprising highly entangled team members and more peripheral ones. Teams might benefit from strong leadership of few selected individuals that can guide and inspire others.

With regard to employees disengagement, other studies have already shown that communication-based metrics of SNA can support the prediction of voluntary turnover (de Oliveira et al., 2019;

Gloor et al., 2017b). We have proven that our proposed metric entanglement can also predict individual employee turnover and might help such studies to improve their model quality.

Secondly, we show that the Gini coefficient of betweenness entanglement, as well as betweenness centrality, are associated with individual employee performance. A high Gini index of betweenness entanglement significantly increases the chance of being a top performer. This means that focused communication – communicating intensively and highly synchronized with a few select colleagues, while reducing communication with the rest of the organization – is an indicator of high performance. Our findings are consistent with past research (Brass, 1984; Mehra et al., 2001; Sparrowe et al., 2001) showing that network centrality is positively related to individual performance. However, the important part of our metric is that synchronization with others has a positive impact on individual performance, and not only having central social position. Centrality alone may not be enough to explain individual performance (Reinholt et al., 2011) and we address this issue with the betweenness entanglement metric. Furtherly, we found that low tenure also has a positive influence on individual performance.

Thirdly, inequality of group betweenness entanglement in teams positively influences customer satisfaction. The company in case study D considers customer satisfaction as a proxy for team performance. Our findings suggest that the stronger leaders with high entanglement emerge in groups, the happier the customer is. This means we have strongly entangled leaders who influence team dynamics over time, while the rest of the team is rather passive. While Mukherjee (2016) reveals a positive relationship between centralized leadership and sport teams' performance, Mehra et al. (2006) suggests that distributed leadership structures can differ with regard to important structural characteristics, and these differences can have positive or negative effects (Cummings and Cross, 2003) on team performance.

This study contributes to research and practice. First, we contribute to synchronization and flow state research by providing a novel metric for determining communication synchronization in working environments. We validated this metric through four case studies in different business contexts. Flow state research can use our metrics to measure team flow not only by conducting surveys but also by using SNA and taking communication into account. Further, we contribute to human resources (HR) research in providing a novel metric for analyzing employee communication and behaviors. Employee communication is a critical factor for good collaboration and employee and team performance (Gloor et al., 2020; Wen et al., 2019), thus using a new metric of communication dynamics might open new research opportunities. On the other hand, decision makers, such as HR managers, could act as interventionist (Valente, 2012) and use this metric to identify weak and strong entangled actors in the communication network of a team or in the entire company. Thus, HR managers might use this metric to improve performance appraisal systems, anticipate disengagement and improve hiring and retention strategies. Combining novel metrics of e-mail communication analysis with long-established methods to assess employees' satisfaction (like surveys), HR managers can offer improved organizational initiatives, such as mentoring programs or cross-staffing, or retention strategies. The entanglement metric described in this paper has the potential to help managers to better understand the nature of employee online communication at their particular organization. This might lead to a rethinking of team design and building in the specific organization, which could ultimately lead to improved communication and collaboration and might support the identification of cohesive groups.

Nevertheless, e-mail communication analysis combined with SNA raises some ethical concerns. HR managers need to make sure that metrics gathered from such analysis are seen as a support for HR decision making, and not as the holy grail for automated decision making (ADM), without questioning the analysis results. False positives or false negatives can occur and emphasize the supportive character of our metric for HR decisions. The goal of our analytical approach is to support general improvement of group performance and employees' wellbeing, also through the recovery of low-performers. There is potential value for senior leaders in monitoring aggregate behaviors, to understand if there are possible waves of disengagement and address them early at the organizational (if not individual) level. In addition, the entanglement metric offers an opportunity for virtual mirroring sessions (Gloor et al., 2017a), where groups and individuals have the chance to self-reflect on their virtual interactions and communication styles. Through virtual mirroring, employees become aware of their online behavior and this usually triggers change, leading to an improvement in communication and performance ("A Novel Way to Boost Client Satisfaction," 2019).

Future studies might also take body measures like heart rate or body movement into account to determine synchronization and flow state during real-time communication (e.g., online/offline team meeting). Nowadays, it is technically easy to collect such data e.g. via smart wearables (Gloor et al., 2018). However, we are aware of the difficulty to use such smart devices in an organizational setting, because of security, privacy, and legal issues. Besides e-mail, employees increasingly use instant messaging tools like Slack or Microsoft Teams. Such tools provide application programming interfaces (APIs) for accessing communication data. Researchers could use that data to build a communication network and follow the analytical approach we presented. In general, our conceptualization of entanglement could be extended to other network measures – such as group betweenness centrality as formalized by Everett and Borgatti (Everett and Borgatti, 1999) – or to other aspects of social interaction – such as measuring synchronicity in the emotions of people who carry out similar activities. In future research, we additionally plan to compare our results with those of other possible approaches to the study of temporal networks (Falzon et al., 2018; Holme and Saramäki, 2012).

Our study has some limitations that should be taken into account. While the evidence supports guidance for new research agendas, our analysis is limited to the contexts of the case studies and the available datasets. It will be important to replicate our analysis in organizations of different industries, also considering different job descriptions and hierarchical positions of employees. We included the entanglement calculation in the SNA tools Condor and Griffin, which are free to use for academics in order to facilitate replicability. Further, other social network metrics could be considered to extend our definition of entanglement. For example, additional interaction patterns could be taken into account, developing metrics that specifically look at who communicates with whom. This could be particularly relevant when additional information about nodes is available, other than the social network structure. In addition, we advocate future studies to more deeply investigate the relationship of entanglement with other social network metrics, both time-variant and time-invariant.

Building upon existing synchronization and flow state literature from different disciplines, we showed that the idea of synchronization and flow state can be used together to develop new metrics – based on methods and tools of SNA. Note that social entanglement is an indicator of behavior, with no definitive claims about cause and causality. Just as with quantum entanglement, much more research will be needed to fully "untangle" the origin of social entanglement. Nevertheless, the findings from our four case studies give evidence to the potential of our proposed entanglement metric. We position our research as a starting point for further HR-related analyses, which consider employees' social interactions and communication, with the goal to improve and optimize collaboration, leading to more satisfied employees and customers.

References

A Novel Way to Boost Client Satisfaction, 2019. . Harv. Bus. Rev. 17–21.

Angier, N., 2018. You Share Everything With Your Bestie. Even Brain Waves. New York Times.

Armstrong, A.C., 2008. The fragility of group flow: The experiences of two small groups in a middle school mathematics classroom. J. Math. Behav. 27, 101–115. https://doi.org/10.1016/j.jmathb.2008.08.001

Arrow, H., Poole, M.S., Henry, K.B., Wheelan, S., Moreland, R., 2004. Time, Change, and Development. Small Gr. Res. 35, 73–105. https://doi.org/10.1177/1046496403259757

Bakker, A.B., 2005. Flow among music teachers and their students: The crossover of peak experiences. J. Vocat. Behav. 66, 26–44. https://doi.org/10.1016/j.jvb.2003.11.001

Brandes, U., 2001. A Faster Algorithm for Betweenness Centrality. J. Math. Sociol. 25, 163–177.

Brass, D.J., 1984. Being in the Right Place: A Structural Analysis of Individual Influence in an Organization. Adm. Sci. Q. 29, 518. https://doi.org/10.2307/2392937

Burt, R.S., 2005. Brokerage and closure. An introduction to social capital. Oxford University Press.

Ceja, L., Navarro, J., 2012. 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. Hum. Relations 65, 1101–1127. https://doi.org/10.1177/0018726712447116

Collins, R., 2005. Interaction Ritual Chains. Princeton University Press.

Condon, W., 1986. Communication: Rhythm and structure, in: Evans, J., Clyne, M. (Eds.), Rhythm in Psychological, Linguistic and Musical Processes. Charles C. Thomas, Springfield, IL, pp. 55–78.

Cramton, C.D., 2001. The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration. Organ. Sci. 12, 346–371.

Cross, L., Turgeon, M., Atherton, G., 2019. How Moving Together Binds Us Together: The Social Consequences of Interpersonal Entrainment and Group Processes. Open Psychol. 1, 273–302. https://doi.org/10.1515/psych-2018-0018

Csikszentmihalyi, M., 2000. Beyond boredom and anxiety. Jossey-Bass, San Francisco, CA, USA.

Csikszentmihalyi, M., 1996. Creativity: flow and the psychology of discovery and invention. Harper Perennial, New York, New York, USA.

Cummings, J.N., Cross, R., 2003. Structural properties of work groups and their consequences for performance. Soc. Networks 25, 197–210. https://doi.org/10.1016/S0378-8733(02)00049-7

de Oliveira, J.M., Gloor, P.A., 2018. GalaxyScope: Finding the "Truth of Tribes" on Social Media, in: Collaborative Innovation Networks. Springer, Cham, pp. 153–164. https://doi.org/10.1007/978-3-319-74295-3_13

de Oliveira, J.M., Zylka, M.P., Gloor, P.A., Joshi, T., 2019. Mirror, Mirror on the Wall, Who Is Leaving of Them All: Predictions for Employee Turnover with Gated Recurrent Neural Networks, in: Collaborative Innovation Networks. Springer, Cham, pp. 43–59. https://doi.org/10.1007/978-3-030-17238-1_2

Delarue, A., Van Hootegem, G., Procter, S., Burridge, M., 2008. Teamworking and organizational performance: A review of survey-based research. Int. J. Manag. Rev. 10, 127–148. https://doi.org/10.1111/j.1468-2370.2007.00227.x

Dubitzky, W., Granzow, M., Berrar, D., 2007. Fundamentals of Data Mining in Genomics and Proteomics, Fundamentals of Data Mining in Genomics and Proteomics. Springer US, Boston, MA, NY. https://doi.org/10.1007/978-0-387-47509-7

Durkheim, É., 2008. The Elementary Forms of Religious Life. Dover Publications Inc.

Einstein, A., Podolsky, B., Rosen, N., 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? Phys. Rev. 47, 777–780. https://doi.org/10.1103/PhysRev.47.777

Elkins, A.N., Muth, E.R., Hoover, A.W., Walker, A.D., Carpenter, T.L., Switzer, F.S., 2009. Physiological compliance and team performance. Appl. Ergon. 40, 997–1003. https://doi.org/10.1016/j.apergo.2009.02.002

Everett, M.G., Borgatti, S.P., 1999. The Centrality of Groups and Classes. J. Math. Sociol. 23, 181–201.

Fairhurst, M.T., Janata, P., Keller, P.E., 2013. Being and Feeling in Sync with an Adaptive Virtual Partner: Brain Mechanisms Underlying Dynamic Cooperativity. Cereb. Cortex 23, 2592–2600. https://doi.org/10.1093/cercor/bhs243

Falzon, L., Quintane, E., Dunn, J., Robins, G., 2018. Embedding time in positions: Temporal measures of centrality for social network analysis. Soc. Networks 54, 168–178. https://doi.org/10.1016/j.socnet.2018.02.002

Freeman, L.C., 1978. Centrality in Social Networks: Conceptual Clarification. Soc. Networks 1, 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Freeman, L.C., 1977. A Set of Measures of Centrality Based on Betweenness. Sociometry 40, 35–41.

Gipson, C.L., Gorman, J.C., Hessler, E.E., 2016. Top-down (Prior Knowledge) and Bottom-up (Perceptual Modality) Influences on Spontaneous Interpersonal Synchronization. Nonlinear Dynamics. Psychol. Life Sci. 20, 193–222.

Gloor, P.A., 2017. Sociometrics and Human Relationships: Analyzing Social Networks to Manage Brands, Predict Trends, and Improve Organizational Performance. Emerald Publishing Limited, London, UK.

Gloor, P.A., Almozlino, A., Inbar, O., Lo, W., Provost, S., 2014. Measuring Team Creativity Through Longitudinal Social Signals.

Gloor, P.A., Colladon, A.F., Grippa, F., Budner, P., Eirich, J., 2018. Aristotle Said "Happiness is a State of Activity" — Predicting Mood Through Body Sensing with Smartwatches. J. Syst. Sci. Syst. Eng. 27, 586–612. https://doi.org/10.1007/s11518-018-5383-7

Gloor, P.A., Fronzetti Colladon, A., Giacomelli, G., Saran, T., Grippa, F., 2017a. The impact of virtual mirroring on customer satisfaction. J. Bus. Res. 75, 67–76. https://doi.org/10.1016/j.jbusres.2017.02.010

Gloor, P.A., Fronzetti Colladon, A., Grippa, F., 2020. The digital footprint of innovators: using email to detect the most creative people in your organization. J. Bus. Res. in press.

Gloor, P.A., Fronzetti Colladon, A., Grippa, F., Giacomelli, G., 2017b. Forecasting managerial turnover through e-mail based social network analysis. Comput. Human Behav. 71, 343–352. https://doi.org/10.1016/j.chb.2017.02.017

Gloor, P.A., Oster, D., Fischbach, K., 2012. JazzFlow—Analyzing "Group Flow" Among Jazz Musicians Through "Honest Signals." KI - Künstliche Intelligenz 27, 37–43. https://doi.org/10.1007/s13218-012-0230-3

Guastello, S.J., Peressini, A.F., 2017. Development of a Synchronization Coefficient for Biosocial Interactions in Groups and Teams. Small Gr. Res. 48, 3–33. https://doi.org/10.1177/1046496416675225

Hancock, J.I., Allen, D.G., Bosco, F.A., McDaniel, K.R., Pierce, C.A., 2013. Meta-Analytic Review of Employee Turnover as a Predictor of Firm Performance. J. Manage. 39, 573–603. https://doi.org/10.1177/0149206311424943

Heyne, K., Pavlas, D., Salas, E., 2011. An Investigation on the Effects of Flow State on Team Process and Outcomes, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 475–479. https://doi.org/10.1177/1071181311551098

Hinds, P.J., Bailey, D.E., 2003. Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams. Organ. Sci. 14, 615–632. https://doi.org/10.1287/orsc.14.6.615.24872

Hinds, P.J., Retelny, D., Cramton, C., 2015. In the Flow, Being Heard, and Having Opportunities, in: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15. ACM Press, New York, New York, USA, pp. 864–875. https://doi.org/10.1145/2675133.2675199

Hoffman, L., Rovine, M.J., 2007. Multilevel models for the experimental psychologist: Foundations and

illustrative examples. Behav. Res. Methods 39, 101–117. https://doi.org/10.3758/BF03192848
Holme, P., Saramäki, J., 2012. Temporal networks. Phys. Rep. 519, 97–125. https://doi.org/10.1016/j.physrep.2012.03.001
Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K., 2009. Quantum entanglement. Rev. Mod. Phys. 81, 865–942. https://doi.org/10.1103/RevModPhys.81.865
Hou, Y., Song, B., Hu, Yinying, Pan, Y., Hu, Yi, 2020. The averaged inter-brain coherence between the audience and a violinist predicts the popularity of violin performance. Neuroimage 211, 116655. https://doi.org/10.1016/j.neuroimage.2020.116655
Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J. Hydrol. 574, 1029–1041. https://doi.org/10.1016/j.jhydrol.2019.04.085
Liu, W., Branigan, H.P., Zheng, L., Long, Y., Bai, X., Li, K., Zhao, H., Zhou, S., Pickering, M.J., Lu, C., 2019. Shared neural representations of syntax during online dyadic communication. Neuroimage 198, 63–72. https://doi.org/10.1016/j.neuroimage.2019.05.035
Lundberg, S., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9
Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, pp. 4768–4777.
Lundberg, S., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., Lee, S.-I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2, 749–760. https://doi.org/10.1038/s41551-018-0304-0
Maznevski, M.L., Chudoba, K.M., 2000. Bridging Space Over Time: Global Virtual Team Dynamics and Effectiveness. Organ. Sci. 11, 473–492. https://doi.org/10.1287/orsc.11.5.473.15200
Mehra, A., Kilduff, M., Brass, D.J., 2001. The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance. Adm. Sci. Q. 46, 121–146.
Mehra, A., Smith, B.R., Dixon, A.L., Robertson, B., 2006. Distributed leadership in teams: The network of leadership perceptions and team performance. Leadersh. Q. 17, 232–245. https://doi.org/10.1016/j.leaqua.2006.02.003
Merten, F., Gloor, P., 2010. Too Much E-Mail Decreases Job Satisfaction. Procedia - Soc. Behav. Sci. 2, 6457–6465. https://doi.org/10.1016/j.sbspro.2010.04.055
Mukherjee, S., 2016. Leadership network and team performance in interactive contests. Soc. Networks 47, 85–92. https://doi.org/10.1016/j.socnet.2016.05.003
Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T., Barabási, A.-L., 2000. The sound of many hands clapping. Nature 403, 849–850. https://doi.org/10.1038/35002660
Nezlek, J.B., 2008. An Introduction to Multilevel Modeling for Social and Personality Psychology. Soc. Personal. Psychol. Compass 2, 842–860. https://doi.org/10.1111/j.1751-9004.2007.00059.x
Parkinson, C., Kleinbaum, A.M., Wheatley, T., 2018. Similar neural responses predict friendship. Nat. Commun. 9, 332. https://doi.org/10.1038/s41467-017-02722-7
Patrashkova-Volzdoska, R.R., McComb, S.A., Green, S.G., Compton, W.D., 2003. Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. IEEE Trans. Eng. Manag. 50, 262–269. https://doi.org/10.1109/TEM.2003.817298
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. Catboost: Unbiased boosting with categorical features, in: NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing. San Diego, CA, pp. 6639–6649. https://doi.org/10.5555/3327757.3327770

Pusey, M.F., Barrett, J., Rudolph, T., 2012. On the reality of the quantum state. Nat. Phys. 8, 475–478. https://doi.org/10.1038/nphys2309

Quinn, R.W., 2005. Flow in Knowledge Work: High Performance Experience in the Design of National Security Technology. Adm. Sci. Q. 50, 610–641. https://doi.org/10.2189/asqu.50.4.610

Reinholt, M., Pedersen, T., Foss, N.J., 2011. Why a Central Network Position Isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. Acad. Manag. J. 54, 1277–1297. https://doi.org/10.5465/amj.2009.0007

Reinke, K., Chamorro-Premuzic, T., 2014. When email use gets out of control: Understanding the relationship between personality and email overload and their impact on burnout and work engagement. Comput. Human Behav. 36, 502–509. https://doi.org/10.1016/j.chb.2014.03.075

Sawyer, K.R., 2003. Group Creativity: Music, Theater, Collaboration. Psychology Press.

Singer, J.D., Willett, J.B., 2009. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence, Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York, NY. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Sparrowe, R.T.., Liden, R.C.., Wayne, S.J.., Kraimer, M.L., 2001. Social Networks and the Performance of Individuals and Groups. Acad. Manag. J. 44, 316–325. https://doi.org/10.5465/3069458

Stevens, R., Amazeen, P., Likens, A., 2013. The organizational neurodynamics of teams. Nonlinear Dynamics. Psychol. Life Sci. 17, 67–86.

Valente, T.W., 2012. Network Interventions. Science (80-.). 337, 49–53. https://doi.org/10.1126/science.1217330

van den Hout, J.J.J., Davis, O.C., Weggeman, M.C.D.P., 2018. The Conceptualization of Team Flow. J. Psychol. 152, 388–423. https://doi.org/10.1080/00223980.2018.1449729

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

Wen, Q., Gloor, P., Fronzetti Colladon, A., Tickoo, P., Joshi, T., 2019. Finding top performers through email patterns analysis. J. Inf. Sci. in press. https://doi.org/10.1177/0165551519849519

Wiltermuth, S.S., Heath, C., 2009. Synchrony and Cooperation. Psychol. Sci. 20, 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x

Yun, K., Watanabe, K., Shimojo, S., 2012. Interpersonal body and neural synchronization as a marker of implicit social interaction. Sci. Rep. 2, 959. https://doi.org/10.1038/srep00959

Zylka, M.P., Fischbach, K., 2017. Turning the Spotlight on the Consequences of Individual IT Turnover. ACM SIGMIS Database DATABASE Adv. Inf. Syst. 48, 52–78. https://doi.org/10.1145/3084179.3084185

Publications

Journal articles

Gloor, P. A., Zylka, M. P., Fronzetti Colladon, A., Makai, M. (2022). ‘Entanglement’ – a new dynamic metric to measure team flow. Social Networks, Vol.70, July, pp.100-111, doi: https://doi.org/10.1016/j.socnet.2021.11.010

Schreiber, R.R., Zylka M.P. (2020) Social Network Analysis in Software Development Projects: A Systematic Literature Review. International Journal of Software Engineering and Knowledge Engineering, Vol.30, No. 3

Zylka, M. P., Fischbach, K. (2017). Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda. ACM SIGMIS DATABASE, Vol.48, No. 2, pp. 52-78, doi: 10.1145/3084179.3084185

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., Schoder, D. (2016). Business model development, founders’ social capital, and the success of early-stage internet start-ups: a mixed-method study. Information Systems Journal (ISJ), Vol. 26, No. 5, pp.421-449, doi: 10.1111/isj.12073.

Posegga, O., Zylka, M. P., Putzke, J., Fischbach, K., Schoder, D. (2014). Understanding the Lifestyle of Health and Sustainability - An Exploratory Study. International Journal of Organisational Design and Engineering (IJODE), Vol. 3, Nos. 3/4, pp.338-357, doi: 10.1504/IJODE.2014.065071

Books

Zylka, M. P., Fuehres, H., Fronzetti Colladon, A., Gloor, P. A. (Eds.) (2016). Designing Networks for Innovation and Improvisation. Springer Proceedings in Complexity.

Book chapters

de Oliveira J.M., Zylka M.P., Gloor P.A., Joshi T. (2019) Mirror, Mirror on the Wall, Who Is Leaving of Them All: Predictions for Employee Turnover with Gated Recurrent Neural Networks. In: Song Y., Grippa F., Gloor P., Leitão J. (eds) Studies on Entrepreneurship, Structural Change and Industrial Dynamics. Springer, Cham

Yang, S., Zylka, M. P., Gloor, P. A. (2018). “German Association or Chinese Emperor?” Building COINs Between China and Germany. Collaborative Innovation Networks - Building Adaptive and

Resilient Organizations. Studies on Entrepreneurship, Structural Change and Industrial Dynamics, Springer, Cham.

Herttua, T., Jakob, E., Nave, S., Gupta, R., Zylka, M. P. (2016). Growth Hacking: Exploring the Meaning of an Internet-born Digital Marketing Buzzword. Designing Networks for Innovation and Improvisation. Springer Proceedings in Complexity.

Gieck, R., Kinnunen, H., Li, Y., Moghaddam, M., Pradel, F., Gloor, P. A., Paasivaara, M., Zylka, M. P. (2016). Cultural Differences in the Understanding of History on Wikipedia. Designing Networks for Innovation and Improvisation. Springer Proceedings in Complexity.

Conference articles/presentations

Zylka, M. P. (2016). Putting the Consequences of IT Turnover on the Map - A Review and Call for Research. Proceedings of the 2016 ACM SIGMIS CPR Conference, Washington, D.C., USA. Best Paper Nominee, doi: 10.1145/2890602.2890618

Posegga, O., Zylka, M. P., Fischbach, K. (2015). Collective Dynamics of Crowdfunding Networks. Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS), Kauai (HI), doi: 10.1109/HICSS.2015.394

Spiegel, O., Abbassi, P., Zylka, M. P., Posegga, O., Schlagwein, D., Fischbach, K., Schoder, D. (2014). Getting Boundary Conditions Right: Towards a Classification of the Information Economy. Academy of Management Proceedings, presented at Academy of Management Annual Meeting, Philadelphia, PA, USA, 1 - 5 August 2014, doi: 10.5465/AMBPP.2014.15984abstract

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., Schoder, D. (2013). The Impact of Structural Embeddedness on Funding Success of Early Stage Web Startups. DIBME 2013 Pre-ECIS Workshop on the Digitization in Business Models and Entrepreneurship, Utrecht, The Netherlands.

Zylka, M. P., Spiegel, O., Fischbach, K. (2013). Employee Turnover in the IT Industry and its Impact on Social Capital and Success of Startup Companies. XXXIII Sunbelt Social Networks Conference of the International Network for Social Network Analysis (INSNA), 21-26 May 2013, Hamburg.