Opening the black box of dyadic knowledge flow

Perceived usefulness of received knowledge in informal knowledge-seeking interactions in a global product development collaboration network

Koudstaal, Gerard

Award date:
2022

Eindhoven University of Technology

Department of Industrial Engineering and Innovation Sciences
Innovation, Technology, Entrepreneurship and Technology

Opening the black box of dyadic knowledge flow

Perceived usefulness of received knowledge in informal knowledge-seeking interactions in a global product development collaboration network

Gerard Koudstaal - 0969039

TU/e Supervisors:
1st: Arjan Markus
2nd: Alex A. Alblas
3rd: Bob Walrave

Company Supervisors:
Jeroen Van Olffen
Mick van der Linden

Eindhoven, February, 2022

Abstract

One of the crucial capabilities of organizations is their ability to use knowledge effectively. However, many organizations struggle to let knowledge flow freely throughout the organization, to link creation and utilization effectively, and increase the productive use of knowledge. With a perspective of friction and redundance, this study aims to provide insight in how firms can enable the flow of knowledge across the dyadic level of knowledge exchange, by investigating knowledge application within a global, cross-departmental knowledge-seeking networks in a product development setting.

Results showed that low tacitness, physical proximity, tie strength, the interaction between cohesion and network diversity, and the receiver not having a managerial role are predictors of successful knowledge application. Tacitness was reduced in cases where the receiver had high network diversity, or where the two individuals in the dyad were both either manager or not a manager. Where they had a different role, it increased tacitness, unless sufficient cohesion was present between the two. For physical distance, knowledge application was only less successful when it concerned inter-continental knowledge exchange. Findings represent important implications for the theoretical framing of dyadic knowledge transfer, the role of physical distance in a digital work environment, and how friction and redundance in network cohesion interact to bring about knowledge application.

Keywords: Dyadic knowledge flow, knowledge application, New Product Development, Knowledge Management, knowledge-seeking collaboration networks

Acknowledgements

I present to you my master thesis “Opening the black box of dyadic knowledge flow: Perceived usefulness of received knowledge in informal knowledge-seeking interactions in a global product development collaboration network”. This thesis represents the result of my thesis project and my internship with ASML Knowledge Management. It marks the end of my master Innovation Management at the University of Technology Eindhoven. I fondly look back on the great time I have had as a student, and am hugely grateful for the people I got to meet along the way. For this project, I would like to thank a couple of those people in particular.

Great thanks go to my supervisors Arjan Markus and Alex Alblas. Arjan, I thank you for being a great sparring partner and an intrinsically interested and engaged supervisor and mentor. I really appreciate the extent to you felt personally invested in my thesis project, and you kept an eye for the personal side of the project. Alex, I thank you for, especially in the second half of my thesis project, basically becoming another first supervisor. You were very involved in the process, and your input improved the empirical quality of my work greatly. I thank you both for your pressure to keep moving forward. Looking back, I really needed that, and have learned a lot about planning and communicating as a result.

I would also like to extensively thank both Jeroen and Mick for their company-mentorship throughout this process. Not only did you enable me to conduct a thesis in an exciting and meaningful context, you provided a team atmosphere, and a truly collaborative project, that I greatly enjoyed. I really appreciated your personal involvement and encouragement. I would also, in extension, like to thank all members of the Knowledge Management as a Service initiative. I have learned a lot participating in planning, preparing workshops, and strategizing about knowledge management, and am really thankful how quickly and completely I was embraced as a part of the team.

I feel I remain in debt to all employees at ASML who participated in my survey. It was them who really made this study possible, and who invested time in such great numbers when the benefit for them was not guaranteed. If I achieved anything in this thesis, I hope it was to impact the way they collaborate and learn together with colleagues in a positive way.

Lastly, I would like to thank friends, family, and my girlfriend, Surja, for their support. They were there to celebrate successes, but also when I had to push myself through a part of the project that was not going the way I expected it to. Much love to all of you.

I wish the reader much joy in reading the result of our collective efforts,

Gerard Koudstaal

Eindhoven, February 2022

Table of Contents

List of Figures

List of Tables

1. Research Problem and Business Relevance

ASML is the largest semiconductor company in the Netherlands, and a global leader in the development, manufacturing and distribution of lithography machines. Their products offer customers the capacity to produce highly advanced integrated circuits. The development of lithography machines on this level is highly complex, and high market demands for development speed forces ASML to go through rapid product improvement and steep learning curves to remain competitive (Case-company, 2021a).

Product development projects are initiated in, and go trough, a product development process. Four departments carry the main responsiblity for designing and developing products. These are named Dept. A, Dept. B, Dept. C and Dept. D. These departments play an important role in making decisions in between new product development stages.

Knowledge management literature, which argues that organizational knowledge has become one of the key assets of businesses (Hislop, 2009). Especially for knowledge-intensive firms like ASML, (Swart & Kinnie, 2003), organizational knowledge is deemed one of its key assets (Nonaka & Takeuchi, 1995; Leonard & Sensiper, 1998). The knowledge based view of the firm poses that the effective integration of individuals' knowledge in goods and services is the primary role of organizations (Grant, 1996), and given that organizations are dynamic, distributed knowledge systems, then "getting connected and interrelating the knowledge each [person] has [...] the key to achieving coordinated action" (Tsoukas, 1996, p.22).

This effective redistribution and application of knowledge is exactly what is strained by organizational silos and barriers, however. As Kleinbaum, Stuart and Tushman (2013) showed, in an unprecedented review of digital communication data between employees, organizational boundaries of units and functions have a significant effect on communication frequency. Functional and departmental boundaries often result in this silo-ing effect, where communication is strained (Allen, James & Gamlen, 2007; McEvily, Soda & Tortoriello, 2014). Especially in innovation, novelty introduces ambiguities that make interpretation across boundaries more arduous (Carlile, 2004). It is especially at these boundaries where fostering knowledge flow can be of crucial importance, as Carlile (2004) concludes that "innovation occurs at the boundaries between specialized domains [and] effectively managing knowledge across the various types of boundaries in an organization is what drives competitive advantage"(Carlile, 2004, p.566). Accordingly, the importance of understanding how knowledge flow is established grows as departmental, functional and geographical boundaries become more prominent, and the potential inhibition of the flow of knowledge gets more severe.

1.1 The usecase

One of the businesslines within ASML is home to many of the company-wide challenges in efficient and effective knowledge flow across organizational silos. ASML knowledge management and Business-line managers are convinced that these issues need to be investigated. Together with people from continuous improvement and quality, they specified a system module, a component of a machine that ASML produces, as a pilot group. It is the ambition of business-line managers to scale findings from this usecase to the other system modules in the business-line. The the usecase is representative of many cross-departmental product development chains within ASML. The group, employing around 250 people across all major departments, presents a highly heterogenous subject group. With three large work-sites, in Europe, The United States and Asia, it covers the three largest work-sites for the businessline. Additionally, the group represents a mix of hard-ware and software, with a roughly 50/50 split in development working on hardware and complementary software.

The cross-departmental usecase consists of twenty-three interdependent teams with the collaborative interdependence of developing, manufacturing, supplying and servicing system module, which is part of several product models and development projects. It is important to note that these groups were implicitly, but not explicitly, defined as being part of the same system module. Nevertheless, these groups are interdependent on product knowledge to collaborate and innovovate together.

The departments and their corresponding teams, are centered around a the stage-gate new product development (NPD) process. Departmental representatives acknowledge the importance of timely availability of the right knowledge within this process. In this view, the effective creation and utilization of organizational knowledge is one of the key drivers of organizational performance, essential for innovation (Jiménez-Jiménez & Sanz-Valle, 2011). Within this NPD-process, key decision moments between product development stages are the place where knowledge on the system module, its past performance, and how it should be improved to enhance customer satisfaction, should come together. At the start of a new product development project, which is run by specified development teams, requirements are needed from other departments, so as to prevent issues down the development pipeline, and prevent mistakes made earlier. As the development project matures, another important interdependence is the handover of knowledge between the departments of how the system module is designed. Departments responsible for manufacturing and maintaining the system module need to be properly informed on the system specifications and inner-workings, to be able to do their work well. The effective application of knowledge is crucial in product innovation processes like these (Cohen & Levinthal, 1990; du Plessis, 2007; Jiménez-Jiménez & Sanz-Valle, 2011).

These knowledge needs are serviced with formal knowledge sharing structures. Key decisions at the end of each stage in the NPD-process have a list of formalized documentation as input and output, which should contain all knowledge needed for requirement integration and effective handovers. This documentation is supplemented with formal meeting structures, with the aim of integrating the required perspectives for key decision making. While the formal NPD-process is a thoroughly matured stage-gate process, it is not, and cannot be, perfect. Stakeholders indicate that handovers are often incomplete or incomprehensible. Documentation is limited or too detailed, and continuous sense-making and perspective taking is required to complement these documents. They also argue that that the right people are not involved from the start, or that people with complementary perspectives in other departments are unknown. As a result, arguably important input is not taken into account at the right moments, or misunderstood, and rejected without consideration. To complement the inevitable shortcomings of formal structures, employees naturally engage in interpersonal discussion and iteration. In an instance in manufacturing where representatives indicated alignment and sharing was going well, they elaborated that at least three cross-departmental alignment meetings a week were required to facilitate the perspective-taking and understanding required to work together.

Busisness-line management and supporting staff have currently initiated an investigation in formal and managerial exchange processes within the usecase. Formal structures do not represent the totality of organizational structures that facilitate knowledge exchange, however. As Allen et al. (2007) put it, "formal channels of communication rarely accurately reflect the working relationships between individuals [...], the myriad of personal communications and ties which in reality disseminate knowledge and information between individuals constitute the informal social networks" (Allen et al., 2007, p.180). McEvily et al. (2014) contribute to this point, adding that "the intricate interplay between formal and informal elements is [...] what ultimately determines individuals' ability to get things done and, consequently, their capacity to facilitate (or, at times, hinder) the pursuit of organizational objectives". (McEvily et al., 2014, p.333). Informal relations form a unique conduit over which knowledge flows, as "informal ties promote vicarious learning. Informal connections allow people to benefit from knowledge accumulated by close contacts" (Argote, McEvily & Reagans, 2003, p.576).

The role of organizational collaboration networks in the diffusion of information and knowledge has been one of the central themes in social network theory since its inception (Powell, Staw & Cummings, 1990; Borgatti & Foster, 2003a). Theories on network closure (Coleman, 1988; Burt, 2004) brokerage (Burt, 2004), boundary spanning (Tushman, 1977), embeddedness (Uzzi, 1996) social capital (Coleman, 1988) all conceive interpersonal network structures to be important in the access to and flow of information resources. Knowledge management literature has acknowledged the importance of relational aspects of knowledge for knowledge sharing, and recognizes that the active context of learning takes place in networks of members, tasks and tools (Argote & Ingram, 2000; Argote & Miron-Spektor, 2011).

The important role of social networks is recognized by employees as well, as ASML is often called a 'network organization'. In the use-case, this interaction is driven by informal knowledge seeking and sharing to elaborate on the content, or to supply information that is not present, or cannot be found, or cannot be written down in digital knowledge repositories. In their day-to-day work, engineers of all departments also need each-others' expertise to resolve issues as they come up.

While revision of formal structures is important, without a good view on informal structures, no holistic set or interventions can be designed, and informal blind-spots might cause omission of relevant organizational aspects, or over-treatment of a knowledge gap in formal structure that social networks already effectively close. As a result, ASML Knowledge Management, who is supporting the initiative, has argued that the workshop on formal and managerial knowledge should be complemented by a review of informal knowledge seeking interactions.

1.2 Informal knowledge flows within the usecase

While stakeholders of the system module recognize that these sharing networks can be very effective at diffusing knowledge across department, there are several contextual challenges that impede the flow of knowledge. Literature recognizes that, while networks themselves are structures of interpersonal interactions (Borgatti & Foster, 2003b), contextual factors have an effect how how effectively they diffuse knowledge (Szulanski, 1996; Ghosh & Rosenkopf, 2015). One of these characteristics is relational, and concerns in what ways which two people engaged in knowledge exchange relate to each other (Szulanski, 1996; Levin & Cross, 2004; McPherson, Smith-Lovin & Cook, 2001; Boschma, 2005; Reagans & McEvily, 2003). Another one is structural, and describes how the personal network is embedded within the complete network of interactions (Reagans & McEvily, 2003; Zhao & Anand, 2013; Kim & Anand, 2018). The last characteristic is the knowledge being diffused, and to what extend it is hard to share and interpret by others (Hansen, 1999; Zander & Kogut, 1995; Sorenson, Rivkin & Fleming, 2006). An investigation into the informal knowledge flow characteristics of the usecase revealed that relational, structural and knowledge characteristic factors played a role in how effective and efficient knowledge could flow between people, across team-, department- and location boundaries.

Relational aspects mostly related to distance, both cognitive as well as physical. As a representative sourcing and supply chain indicated, geographic barriers made transfer and alignment cost more time. “Out of sight, out of mind” - is a way an interviewee of Dept. B in the U.S.A. described this difficulty. A representative of Dept. A remarked that “when you spread teams across different parts of the world, with time difference, different cultures, it has to be managed. [...] it requires more alignment”. “The further away, the larger the problem of course”, a representative of Dept. B in Europe remarked about the collaboration with Asia. Proximity literature has shown that these effects can influence the effective flow of knowledge, arguing that close proximity exposes actors to positive externalities, and facilitates informal relationships (Boschma, 2005).

A representative of Dept. A indicated that common knowledge is needed to understand eachother. “Often”, a Dept. B representative elaborates, “complexity is not even relevant, because you do not even get to that level”. As a result, “80% of our time is spent understanding eachother”, says a Dept. B representative, responsible for integration cross-departmental product requirements. The degree to which people are able to understand each-others perspective through common perspectives and knowledge, also termed cognitive distance, has been identified as an important relational factor when it comes to the effective flow of knowledge. Theory on cognitive distance shows that common knowledge is required to be able to integrate knowledge (Cohen & Levinthal, 1990), especially in such an interdisciplinary network.

Knowledge characteristics were also deemed an influential factor by interviewees. ASML is home to a high concentration of very specialized and highly educated people, to work on advanced, technologically complex systems (Case-company, 2021a). A recent study on the employee experience of knowledge management found that complexity plays an important role seeking and sharing (Case-company, 2021b). A representative of Dept. B in the U.S.A. remarked that “You cannot completely hide complexity”, and an interviewee of Dept. D in the U.S.A. stated that complexity can be an issue when sharing takes place between people with a difference in educational background. Complexity has indeed been shown to play a role in the successful flow of knowledge, especially between organizational units (Hansen, 1999; Hansen & Løvås, 2004), like in the use-case.

While the role of structural characteristics were less evident from interviews, having 'the right network' was something interviewees stressed was important. A Dept B. manager, who recently moved to the business-line, underlined the necessity of knowing the right people, indicating that he was dependent on others to connect him to knowledge he needed. In contrast, the inability to connect the "right people" to meetings and discussions was seen as a barrier to improving decision quality.

It is widely known that these contextual factors play an important role in the diffusion of knowledge. Not enough is known, however, *to what extent* the contextual factors of the use-case impact the ability of employees within the optical column usecase to exchange knowledge effectively. A clear image of how these contextual factors obstruct and influence the effectiveness of knowledge flows, and the successful utilization of cross-departmental knowledge between departments, remains elusive.

1.3 Research questions

An understanding of now informal networks complement formal structures of knowledge transfer is crucial for managers of the use-case in their effort to improve cross-departmental knowledge flows. A systematic study of the structure of the informal networks, and to what extent contextual factors interact with the ability of the network to effectively disseminate knowledge, is required. A key component to built this understanding is to identify how the dyadic transfer relationships lead to the successful knowledge application, that creates business impact. Since there is interest in ASML to increase the use of network analysis as a tool to understand and improve knowledge flows, the need for the development of an appropriate methodology to study knowledge flows, that can be applied elsewhere, provides further value to this study. For this study, the receipt of useful knowledge (Levin & Cross, 2004) is used to model knowledge application in dyadic knowledge transfer. This is further elaborated in the theoretical background.

Following the case description and the context provided by stakeholders and managers in the NPD-network, the research question of this report is as follows:

> *"To what extent do relational characteristics and knowledge characteristics impact the relation between structural network characteristics and the receipt of useful knowledge from dyadic knowledge flow, for the collaborative network of the usecase?"*

In order to answer this research question, the following sub-questions are formulated:

- How do relational, structural and knowledge characteristics interact with the receipt of useful knowledge flow from dyadic knowledge flow in the usecase?
- In what way can knowledge flow within the usecase network be improved, based on current knowledge flows, and related to innovative performance?

This study will present answers to these questions in the following manner. First, the theoretical background is established in chapter 2, providing a theoretical perspective with which to address the questions raised in the introduction. Subsequently, in chapter 3, a description of the methodology is provided, detailing the survey study conducted to investigate dyadic knowledge flow in the usecase. Afterward, in chapter 4, results of this study are presented. Finally, in chapter 5, conclusions are drawn, and findings are discussed, and related to literature.

This thesis also presents appendices. Most importantly, appendix A, presents information relevant to the context of this study, but removed for confidentiality reasons. Review this appendix for a non-redacted introduction, more network visualizations, interaction occurance heatmaps and more context specific recommendations.

2. Theoretical Background

This chapter describes the theoretical framework adopted for this thesis. The chapter starts in section 2.1 with a discussion on knowledge flows in literature, and explains the focus on dyadic knowledge flow. In section 2.2, dyadic knowledge flow is deconstructed, highlighting stages, characteristics and levels of analysis. Afterward, the adoption of theoretical perspective of friction and redundance is motivated in section 2.3. The following section, section 2.4, argues how this perspective applies to dyadic cohesion. Section 2.5 subsequently discusses moderators relevant to the theoretical model, which is presented and discussed in section 2.5.6. Finally, the chapter ends in section 2.6 by discussing the academic relevance of testing the theoretical model presented in this chapter.

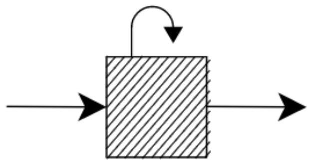
2.1 Conceptualizing knowledge flows

There are two streams of thought when it comes to how knowledge is characterized, which has repercussions on how it can flow (Hislop, 2009). The objectivist perspective sees knowledge as an asset which can be freely shared (Hislop, 2009). This perspective is dominant in early knowledge management literature (Nonaka & Peltokorpi, 2006), and assumes direct transfer between a sender and receiver (Hislop, 2009). The other perspective, called subjective (Nonaka & Peltokorpi, 2006), or practice-based (Hislop, 2009), describes knowledge to be embedded in context, subjective, and therefore harder to transfer. As these two perspectives intrinsically view organizational knowledge differently, these can also lead to two different conceptualizations of how the flow of knowledge should be organized and measured. Objectivist studies are more likely to focus on the collection and codification of knowledge, where subjectivist studies focus more on interpersonal knowledge sharing (Nonaka & Peltokorpi, 2006; Hislop, 2009).

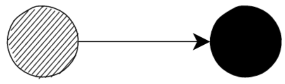
Both knowledge management and organizational learning literature consider the flow of knowledge. Organizational learning concerns itself with subjects related to organizational knowledge creation and acquisition (Castaneda, Manrique & Cuellar, 2018). Knowledge management concerns the deliberate effort to manage knowledge within an organization (Hislop, 2009), focused on the identification, capture, selection, storage, sharing, application, creation and selling (Liebowitz, 1999) of organizational knowledge. There is significant overlap between these fields (Castaneda et al., 2018). As such, both are considered in this study.

Literature of the movement of knowledge within the firm, with its effective utilization as a result, is broad and multifaceted. There are three main streams of literature on knowledge flow, shown in figure 2.1; as stock and flow, as active exchange between two people in a dyad, and as diffusion among a network of interpersonal ties. These different perspectives are elaborated below.

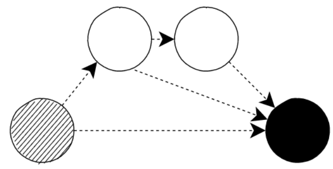
One stream of literature considers flows as “both the stock and flow of knowledge in the organization’s memory” (Argote & Miron-Spektor, 2011, p. 1130), and is part of knowledge retention literature (Argote & Miron-Spektor, 2011). This stream of literature takes the organization as the aggregate level of study as a focal point. It subsequently studies the effective management of its knowledge stock, considering the drivers of knowledge flow into and out of the focal firm to achieve desirable creation and acquisition (Argote & Miron-Spektor, 2011) and organizational unlearning (De Holan & Phillips, 2004).



(a) Flow as a stock and flow model - Knowledge flows can be viewed from the perspective of a stock and flow model of a single entity. In this case, knowledge flow concerns (un)learning (Gupta & Govindarajan, 2000; Schulz, 2001; De Holan & Phillips, 2004; Argote & Miron-Spektor, 2011).



(b) Flow as dyadic exchange or transaction - Knowledge flows can also be studied as a transmission or exchange between two 'entities', in a dyadic interaction, commonly between people, organizational units or firms. This concerns literature on knowledge transfer (Argote & Ingram, 2000), knowledge sharing(Wei et al., 2011) and knowledge exchange; the interaction between two individuals, at different levels of analysis. Can only be dyadic.



(c) Flow as diffusion or dissemination - Knowledge flows can also refer to the study of knowledge movement between three or more 'entities'; in this case, it concerns dissemination, or diffusion, between nodes in a knowledge network. This can also be dyadic, if the origin and the destination of knowledge are investigated, but a multistep diffusion process is implied (for instance citation studies like Sorenson et al. (2006); Ganco (2013)).

Figure 2.1: Knowledge flow in literature

The second stream of literature, which is more extensive, concerns itself with how knowledge and information is exchanged between two actors. These dyadic knowledge flows are closely related to concepts such as transfer, sharing or acquisition, often used interchangeably (Wijk, Jansen & Lyles, 2008), where literature on knowledge transfer focuses on on the interaction or influence between two units (organizations or people) (Argote & Ingram, 2000), and sharing often considers the "provision of task information and know-how" (S. Wang & Noe, 2010, p. 117).

The third to conceptualize knowledge flows is diffusion, and how knowledge moves through and across organizational boundaries between networks of interactions (Sorenson et al., 2006). These flows are the result of a multitude of dyadic transfers or sharing interactions on a network level (Cowan & Jonard, 2004). When actor A transfers knowledge to actor B, and actor B subsequently transferred knowledge to actor C, we cannot say that actor A transferred knowledge to actor C, but knowledge has still flowed between A and C. As such, the concepts of flow, transfer and sharing are considered to be very similar, but the flow of knowledge stretches more broadly across perspectives.

As established, organizations are distributed knowledge systems (Grant, 1996; Tsoukas, 1996), where formal and informal interactions form networks over which knowledge is diffused (Argote et al., 2003; Allen et al., 2007; McEvily et al., 2014). As organizational knowledge resides in its members (Grant, 1996), interpersonal exchange between colleagues is an important part of the creation, diffusion and subsequent application of organizational knowledge (Alavi & Leidner, 2001). In the pursuit of performance in an NPD context, employees are continuously dependent on the exchange of knowledge to recombine and retain organizational knowledge. As a result, dyadic knowledge flow plays a crucial role in the studying of informal knowledge sharing within the usecase.

Although sometimes termed equivalent to transfer and sharing (Wijk et al., 2008), dyadic knowledge flow as it is characterized in transfer and sharing literature is different from the definition used in diffusion studies, where the emphasis lies more on the relocation and reconstruction of knowledge (Rivkin, 2001; Sorenson et al., 2006; Zhao & Anand, 2013). This step, where knowledge needs to be successfully learned and adopted, seems missing from most transfer and sharing studies. This study tries to bridge these distinctly different conceptualizations of knowledge flow. As a result, the distinction of knowledge flow separate from transfer and sharing is in order, and conclusions will be drawn both for our understanding of dyadic exchange and the diffusion of knowledge within organizational product innovation networks. Subsequently, this study defines dyadic knowledge flow as “the movement of knowledge between people, as a result of individual transfer or sharing, in such a way that it has been internalized and can be applied by the receiver”, and models the movement of knowledge from actor to actor through a diffusion network.

2.2 Opening the black box of dyadic knowledge flow

Literature has historically establish numerous ways to measure dyadic knowledge flow. As knowledge is contextual (Argote & Miron-Spektor, 2011), subjective (Hislop, 2009) and embedded in its environment, it must be directly or indirectly approximated or implied, either through behaviours, perceptions or outcomes, especially when knowledge is tacit or ambiguous (Szulanski, 1996). In a review of knowledge transfer antecedents, Filieri and Alguezaui (2014) discuss dependent variables used in literature. This review, together with other works, is used to get an overview of how knowledge flow can be operationalized, and also forms the basis of table 2.1. In knowledge flow literature, studies try to subjectively approximate the actual volume of knowledge flow, either through surveys or in interviews, or studying outcomes.

In their review of knowledge transfer, Filieri and Alguezaui (2014) also define several knowledge transfer processes. These processes were integrated in a broader set of stages adapted to dyadic knowledge flow, presented in table 2.1. The definitions in this study differs slightly, as dyadic knowledge flow also concerns the relocation of knowledge; the internalization to such extent that it can be applied, retained, and shared again by the recipient. This multi-faceted characterization of knowledge flow shows how it is difficult for individual measures of knowledge flow to measure the complete knowledge flow processes. While arguably not all stages need to occur for knowledge to flow, as someone might receive knowledge without having looked for it, or recombine existing elements of knowledge already possessed to created new knowledge, by recombining existing elements, true reconstruction of knowledge at the receiver (Sorenson et al., 2006; Ghosh & Rosenkopf, 2015) can only occur when most stages have succeeded. While most studies characterize their measure of knowledge flow less limited as they are depicted in table 2.1, and imply complete transfer or flow, a thorough review exposes this generalization to scrutiny. As a result, this study discusses knowledge flows, but limits identified implications to knowledge application.

Literature has extensively studied characteristics of dyads that influence the potential of knowledge to flow between them. These fall into four categories (Argote et al., 2003; Levin & Cross, 2004); individual, relational, structural and knowledge characteristics. Individual characteristics concern personal factors that influence motivation and ability (Argote et al., 2003) of individuals to share, transfer and use new knowledge, separate from with whom they interact. Relational characteristics refer to how two people in a dyad relate to each-other; different kinds of proximity, like cognitive or physical proximity (Boschma, 2005), that influence the way they engage in

Stage	Definition	Examples of measures
Knowledge search	The activity of looking for and identifying useful knowledge outside oneself	Tie formation probabilities (Sorenson et al., 2006) [1], knowing what others know and valuing others' knowledge (Borgatti & Foster, 2003a), sought knowledge and search costs (Hansen, Mors & Løvås, 2005) and transactive memory systems (what (Lewis & Herndon, 2011) would refer to as TMS structure)
Knowledge access / exchange	The activity of accessing external knowledge	Most closely related to sharing and transfer in literature. Measures include knowledge transfer behaviour (Kang & Kim, 2010; Wei et al., 2011), ease of knowledge transfer (Reagans & McEvily, 2003) or perceived transfer costs (Borgatti & Foster, 2003a; Hansen et al., 2005), knowledge transfer speed (Zander & Kogut, 1995), interaction frequency (Borgatti & Foster, 2003a), transfer event (Hansen & Løvås, 2004)
Knowledge internalization	The process of analysing, processing, interpreting and understanding the knowledge obtained from external sources to combine it with existing internal knowledge	Related to assimilation, absorption and integration, and the degree to which the receiver can relate the knowledge to what he/she already knows. Measures include knowledge acquisition (Tortoriello, Reagans & McEvily, 2012).
Knowledge application / retention	The activity of continuously applying knowledge obtained from external sources	Knowledge outcomes; use, retention and recombination and creation. Receipt of useful knowledge (Levin & Cross, 2004), creativity (Perry-Smith, 2006), impact factor (McFadyen, Semadeni & Cannella Jr, 2009)

Table 2.1: The stages of dyadic knowledge flow

The first three stages were based on a literature review on knowledge transfer by (Filieri & Alguezaui, 2014).

the exchange of information. Structural characteristics refer to the network structure in which the dyad is embedded, that determines informational access and ability to disseminate (Burt, 2001). Finally, knowledge characteristics refer to the nature of what is being diffused; its tacitness (Nonaka & Takeuchi, 1995), complexity(Zander & Kogut, 1995), or causal ambiguity (Szulanski, 1996).

Not all studies in this field focus on all four characteristics, often only covering one, two or three of the characteristics identified. Even studies that present a holistic perspective, don't consider there to be four characteristics, but three. Argote et al. (2003), for instance, identify three characteristics; individual, relational, and knowledge. Levin and Cross (2004) argue that relational, structural and knowledge related characteristics are relevant. Most studies control for unforeseen characteristics, by designing their methods to control for individual characteristics (Reagans & McEvily, 2003; Levin & Cross, 2004), but this still limits the interpretation of how characteristics interact.

Important in the measurement of knowledge flow is the distinction between the perspectives of sender and receiver. Dyadic knowledge flow happens in a collaboration between two people, but experiences of that dyadic interaction can differ between participants. Generally, receivers evaluate successful knowledge flow, such as knowledge acquisition or application (Tortoriello et al., 2012; Levin & Cross, 2004), as they are the ones implementing knowledge transferred. Senders, on the other hand, can evaluate their unique perspective on transfer costs (Reagans & McEvily, 2003). To get the 'full-picture', of continuous exchange, researchers sometimes rely on managers to evaluate both perspectives as an observer of interactions (Hansen & Løvås, 2004).

Another important aspect of knowledge flow is the level of analysis. As outlined before there are relational and individual characteristics, but these can also both be measured on an individual or dyadic level. Age or job grade, individual characteristics, can be expressed in a dyadic way, by measuring age or job-grade difference (Tortoriello, McEvily & Krackhardt, 2015). Cohesion, a structural measure of dyadic embeddedness, can be measured from an individual perception of cohesion (Levin, Walter, Appleyard & Cross, 2016), or the accumulation of perceptions (Reagans & McEvily, 2003; Tortoriello et al., 2012). As a result, studies can choose to measure all variables dyadically, which is often the case, motivated by homophily theories, but can also be studied in a multi-level model. A model like this incorporates variables measured on the dyadic and individual level, like in (Kang & Kim, 2010).

Since this study is interested in degree of success of the flow of knowledge in a network of interdependent actors, a dependent variable from the receiver, which can evaluate the business value of knowledge flow, is most appropriate. The variable that seems suitable to evaluate this, is a measurement of knowledge application, the 'receipt of useful knowledge', conceived by Levin and Cross (2004). It is a dyadic flow measure, aimed at evaluating dyadic relationships, but the usefulness of the relationship is not measured in whether transfer took place (Hansen & Løvås, 2004), whether mutual understanding was reached (Tortoriello et al., 2012) or whether it was easy to transfer (Reagans & McEvily, 2003), but the result of all three; the impact the knowledge made. While "recipients of knowledge may not always acknowledge when they have acquired new knowledge or accurately identify the source of knowledge" (Reagans & McEvily, 2003, p.243), studies on dyadic knowledge flow rarely combine the perspective of sender and receiver (apparent from the review of Filieri and Alguezaui (2014), presumably because of resource constraints), and the receiver perspective is expected to best estimate the impact of knowledge. Survey measures are related to client satisfaction, team performance and project value, quality and cost efficiency. Since it is ASML's goal for dyadic transfer to create business impact, this is the dependent variable used to model successful knowledge flow.

2.3 Network friction and redundance

Ghosh and Rosenkopf (2015) argue network studies have been too inconsiderate of dyadic transfer frictions in evaluating the effectiveness of a knowledge network. Networks that are similar in structure might have very different patterns of diffusion as a result of dyadic friction and the inability of dyads to transfer, even on short network paths (Ghosh & Rosenkopf, 2015). In the ASML context, this is characterized by several instances where interviewees indicated the application of one an-other's knowledge was just not successful, even if it was attempted. Quotes like "everyone within ASML thinks they know it, but there's few who really understand it, you really need to find those. It depends on the quality of your network" (representative Dept. B) and "we don't get much information [...] we're kind of working in the dark" (representative Dept. D, from Asia) remind us that not every network connection is the same, and not all connections diffuse as effectively. As a representative from Dept. A puts it, "frictions originate from the fact that the implications of adjustments for other departments have to be explained".

Due to how difficult it can be to separate knowledge from individuals, and to motivate and enable individuals to share knowledge effectively, called the stickiness of knowledge by Szulanski (1996), reconstruction and reinterpretation of knowledge is needed at every step in the network (Sorenson et al., 2006). Swan, Newell and Scarbrough (1999) argue that 'knowledge networks' should be conceptualized as networking, putting emphases knowledge transfer as an active process, dependent on "relationships, shared understandings and attitudes" (Swan et al., 1999, p. 273). In the evaluation of a knowledge sharing network, the small world paradigm of clustered networks with short paths between clusters (Watts & Strogatz, 1998; Humphries & Gurney, 2008; Telesford, Joyce, Hayasaka, Burdette & Laurienti, 2011) found optimal in simulations (Cowan & Jonard, 2004) is not enough; understanding how context interacts with the network effectiveness is important, and the understanding of dyadic knowledge flow is therefore crucial to understand larger diffusion patterns in networks. Dyadic knowledge flow frictions can result from structural characteristics (Reagans & McEvily, 2003), knowledge characteristics (Sorenson et al., 2006) and relational characteristics (Levin & Cross, 2004) of networks (Ghosh & Rosenkopf, 2015).

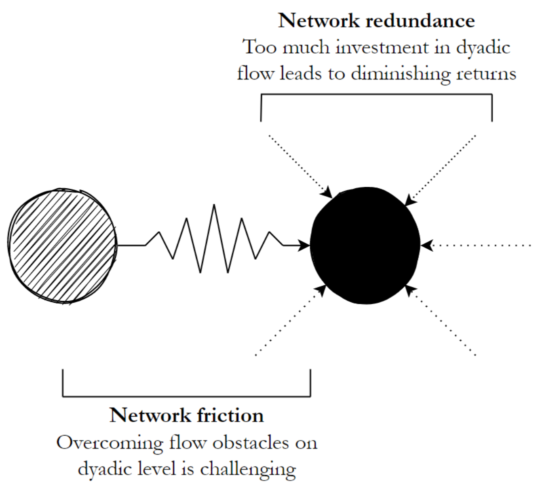


Figure 2.2: Friction and redundance

While a friction view underlines the importance of cohesive structures and network redundancy in alleviating friction (Reagans & McEvily, 2003; Ghosh & Rosenkopf, 2015), the cost of redundancy should not be overlooked. The time investment needed to create cohesive structures and relations can be significant, while it might lead to redundancy in informational access (Burt, 2004; Alguezaui & Filieri, 2010). Logically, with every addition of a network connection, the knowledge overlap between an ego and its network can only increase. Ideally, network redundance should only occur in those places in the network that call for closure to eliminate friction. Highly cost-effective, low redundancy connections can be just as important, if not more, for effective network structures (Granovetter, 1973). Network closure, or cohesive structures of closely connected actors, just like relational redundance, should be framed as a costly tool that can increase effectiveness, but decrease efficiency. This frame is not always adopted in literature. For instance Tortoriello et al. (2012) show cohesive structures enable acquisition of knowledge across organizational units, but to not address the costs of such closure. This report aims to take both a frictional and redundance view on dyadic knowledge flow, and more specifically application, to identify the point of network redundance that alleviates friction, to such an extent that it optimizes the value of dyadic ties, for several contextual factors.

2.4 The central role of cohesion

One of the most fundamental characteristics of a networks' structure is its cohesion. Cohesion, visually represented in figure 2.3, is a measure of triadic closure in a dyad. It describes the degree to which the two people involved in dyadic knowledge flow share common connections in their personal networks. Network closure has been part of one of the longest standing debates in social network literature on the effectiveness of networks to diffuse knowledge (Coleman, 1988; Burt, 2001; Borgatti & Foster, 2003b; Filieri & Alguezaui, 2014), and cohesion has been related to many aspects important to dyadic knowledge flow, such as establishing trust (Coleman, 1988), developing common knowledge (Reagans & McEvily, 2003) and facilitating reciprocity and shared norms (Coleman, 1988). In literature, cohesion is often considered the reverse of network diversity, both in the way it is modeled (Reagans & McEvily, 2003; Levin et al., 2016) as well as the implied benefits, where network diversity exposes and individual or a dyad to new and converging ideas (Burt, 2001; Filieri & Alguezaui, 2014).

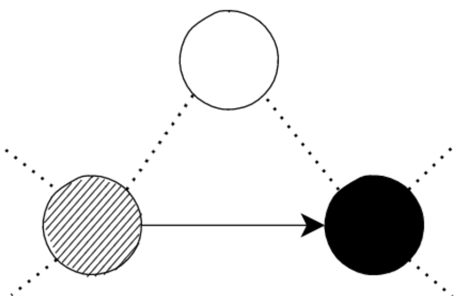


Figure 2.3: Cohesion: triadic closure in dyadic knowledge flow

There are two sides to network cohesion. Firstly, cohesion has been shown to be an important facilitator of the transfer of knowledge between colleagues (Reagans & McEvily, 2003), also across organizational barriers (Tortoriello et al., 2012), alleviating friction (Ghosh & Rosenkopf, 2015).

Recent studies show that cohesive structures have been shown to be needed for successful brokerage, more so than a dyadic exchange alone (Tortoriello & Krackhardt, 2010; Tasselli & Caimo, 2019), and that strong ties might not be strong enough to bridge structural holes. Kim and Anand (2018) propose that cross-unit cohesive structures are especially important when transferring more complex knowledge. As Mors (2010) put it; "[in] heterogeneous contexts, dense network interactions facilitate partners' ability to integrate the diverse information to which they are exposed" (Mors, 2010, p.841).

On the other side, cohesion limits the availability of new perspectives, inducing informational redundance. In an extensive review of networks with high and low cohesion (sparse networks), Alguezaui and Filieri (2010) conclude while networks that are non-cohesive can impede frequent interactions, experience action problems and poor knowledge understanding, cohesive networks can suffer from costly relationships, redundant flows, and restricting novel re-combinations. In the study of (Tortoriello et al., 2015), cohesion is literally called network redundance, and negatively related to innovative performance. Simulation studies propose that in complete cohesion, where *all* ties are shared, there should theoretically be very little novelty to a given connection (Cowan & Jonard, 2004).

Resulting from these two sides is a conflict between redundance and friction that has not completely been resolved by literature. Reagans and McEvily (2003) show that cohesion eases knowledge transfer, but Tortoriello et al. (2015) show cohesion is negatively related to innovative performance, both aspects of knowledge flow. (Levin et al., 2016) remained inconclusive on the role of cohesion, while also employing the receipt of useful knowledge.

An inverted-U shape relationship offers a potential explanation. (Wijk et al., 2008) specifically called for an investigation of curvelineair effects in the antecedents of knowledge flow, and argued that "as the number of relations grows beyond a certain level, the time, energy and attention needed to establish and maintain such relations may diminish knowledge transfer" (Wijk et al., 2008, p.848). This curvilinear effect has been observed for cohesion and knowledge sharing behaviour, for instance (Yu, Hao, Dong & Khalifa, 2013), and embeddedness has shown diminishing returns for firms (Uzzi, 1996). (Kao, Su & Chen, 2019) build on these findings to show that there is diminishing return for interconnectedness in in the efficiency in customer networks. Simulations of cohesion show an inverted U-curve too, where some cohesion is needed to increase the probability of successful transfer (Zhao & Anand, 2013), but too much cohesion is suggested to be inefficient (Cowan & Jonard, 2004), as network redundancies get larger, and no single actor is able to transfer particularly useful knowledge for business outcomes, the benefit of cohesive structures diminishes (Kim & Anand, 2018). A curvelineair effect between cohesion and successful dyadic knowledge flow, such as the receipt of useful knowledge, has not been established, or discussed, by any of the empirical investigations of dyadic knowledge flow studies, however.

Inverted U-shape relationships occur between a dependent and an independent variable when the dependent variable, for instance the receipt of useful knowledge, is caused by two counteracting effects, that both interact with the independent variable, and at least one of these effects is curvilinear (Haans, Pieters & He, 2016). To clarify, this interaction between counteracting effects for cohesion is shown in figure 2.4. This study views these opposing forces to be the removal of friction and the introduction of redundance, which as discussed in section 2.3 both have well-established relationships with cohesion in literature. The linear relationship between cohesion and reduced friction was found by (Reagans & McEvily, 2003), who showed cohesion improved the ease of transfer, and the impact of redundance was found by (Tortoriello et al., 2015), showing a negative impact on innovative performance. Studies that have proposed curvilinear effects have implicitly assumed there to be curvilinear effects in either friction or redundance, in the effects found in survey (Yu et al., 2013) and simulations (Zhao & Anand, 2013; Kim & Anand, 2018). The arguments used revolve around diminishing returns the amount of relationships that need to be managed becomes overbearing (Wijk et al., 2008), cognitive overload increases (Kim & Anand, 2018) and marginal returns on usefulness of knowledge diminish. Studying the relationship between cohesion and knowledge application in dyadic exchange is expected to help address this gap in literature.

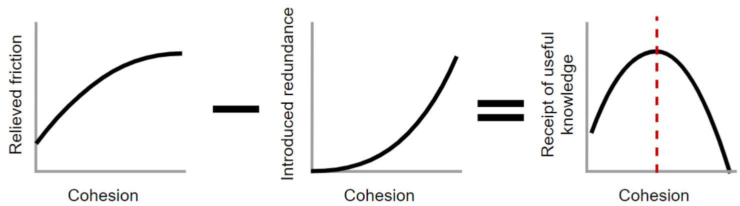


Figure 2.4: The hypothesized curvilinear effect between cohesion and the receipt of useful knowledge
Figure based on visualizations by (Haans et al., 2016). The red line represents the hypothetical optimum point of cohesion.

This study considers cohesion to be both making the transfer of knowledge easier, but also to introduce a diminishing return as the amount of shared relationships inevitably diminishes the usefulness of dyadic transfer. As a result, the following hypothesis is formalized:

Hypothesis 1: Cohesion has an inverted U-shape relationship with the receipt of useful knowledge, such that the quadratic has a significant, negative coefficient

2.5 Moderation effects

Based on the view of friction and redundancy, the optimal level of network cohesion is reasoned to depend on knowledge characteristics, structural characteristics and relational characteristics, and less prevalent in individual characteristics of dyadic knowledge flow. Based on the context, the need for redundancy to reduce friction could be larger, or other relational factors could substitute cohesion to facilitate effective flow. These moderators would influence the impact cohesion has on the reduction of friction, and, hence, on the curvilinear relationship between cohesion and the receipt of useful knowledge. Figure 2.5 shows an example where a contextual factor alleviates a lot of friction in a dyad, reducing the impact cohesion has on successful transfer, making the weight of its redundance more impactful, and lowering the optimum level of both the possible receipt of useful knowledge, and the level of cohesion at which this optimum is reached.

2.5.1 Tie strength

Tie strength, a relational characteristic, is a measure of how intensive the working relationship between two people is. It refers to the frequency in which two people involved in a dyad interact,

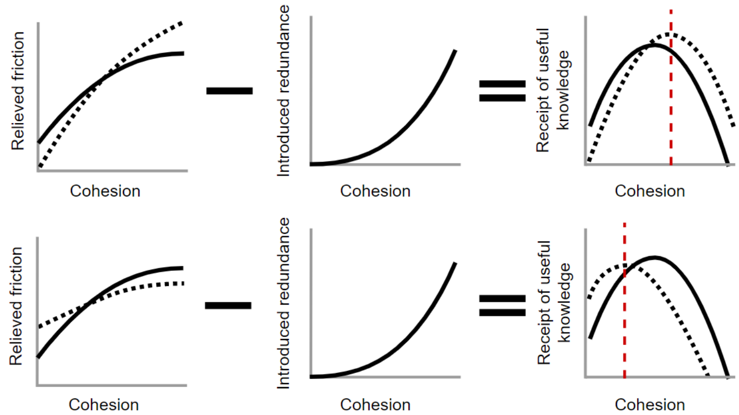


Figure 2.5: The hypothesized effect of moderators on cohesion
Figure based on visualizations by (Haans et al., 2016). Dotted lines show the change of the relation between cohesion and friction through a moderator. The red line shows how the hypothetical optimum point of cohesion moves along the x-axis as moderators increase or decrease optimum point of cohesion for the receipt of useful knowledge.

and to what extend they feel close to each-other (Reagans & McEvily, 2003).

In situation of high tie strength, keeping all other factors constant, cohesion is expected be more redundant, decreasing the optimal level of cohesion for the receipt of useful knowledge. Alternatively, in a situation of low tie strength, more cohesion is expected to be needed to bridge distance between two individuals, increasing the optimal level of cohesion for the receipt of useful knowledge.

Tie strength has been found to be an important enabler for intra-organizational knowledge transfer (Wijk et al., 2008), fostering trust (Levin & Cross, 2004), also across groups (Nakauchi, Washburn & Klein, 2017) and organizational units (Hansen, 1999; Tortoriello et al., 2012). Hansen (1999) showed that weak ties impede the transfer of fine-grained, complex knowledge across organizational units. Levin et al. (2016) argued that tie strength can be a substitute for network cohesion in forming a bridging tie across groups. These studies show that tie strength can close social distance and the larger informational redundance, can be a requirement for successful transfer. Additionally, (Levin et al., 2016) showed that trust, something they closely associated as the driving force behind tie strength (Levin & Cross, 2004), had a moderating effect with cohesion in their study. Tie strength could therefore be a substitute of, or complement to, cohesion in the network.

Hypothesis 2: Tie Strength moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in tie strength lowers the optimal level of network cohesion for receipt of useful knowledge.

2.5.2 Physical distance

Physical distance, another relational characteristic, concerns the distance between the workplaces of the dyad. It relates to how easy it is for two people involved in a dyadic exchange to meet physically to exchange ideas, or to what extend they need to use communication tools to communicate, and the probability of running into each-other on the work-floor. People that work further away form each-other are less exposed to knowledge spill-overs (Boschma, 2005). The common norms and reciprocity (Coleman, 1988), shared language (Haldin-Herrgard, 2000; Tasselli, Zappa & Lomi, 2020) built in cohesive structures (Coleman, 1988), needed for transfer, are also expected in where colleagues are geographically proximate, run into each-other frequently, or establish an office culture. Homophily theory shows that geographic proximity increases the frequency of interaction (McPherson et al., 2001; Kleinbaum et al., 2013). Actor that are physically close share have more in common in general, such as participation in local communities or cultures. More similar relationships, according to McPherson et al. (2001), are more conductive to the receipt of information. Actors in a dyad that work far apart, that despite modern communication tools can communicate with no delay, are expected to experience friction in knowledge flow. With two geographically proximate actors, knowledge should flow more easily, and hence less cohesion is needed to alleviate friction. Furthermore, structural cohesion might be substituted by cohesion in local communities or practices. When two actors are geographically distant, it is expected that a higher level of cohesion is needed to bridge this relational gap, and that cohesion is less redundant. Therefore, it is hypothesized that:

Hypothesis 3: Physical distance moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in distance increases the optimal degree of network cohesion for receipt of useful knowledge.

2.5.3 Range

Range, or network diversity, is a structural characteristic of the two individuals in a dyad. It refers to the extend to which the network of both the ego and the alter in dyadic exchange have networks that provide access to a diverse set of knowledge.

Range has shown to be a conductor of knowledge transfer (Reagans & McEvily, 2003)), also across organizational units (Tortoriello et al., 2012). Range offers access to diverse, non redundant resources, providing brokerage opportunities (Burt, 2004) that should increase the value of transfer regardless of its effort. Additionally, range gives the actors in dyadic relationships the capability to integrate knowledge from different backgrounds (Reagans & McEvily, 2003; Tortoriello et al., 2012). Since actors with higher range are better at effectively using diversity in the network, it is expected that a high level of range makes cohesion more redundant, decreasing the optimal level of cohesion. When two actors have a low range, more cohesion is expected to be needed for the receipt of useful knowledge.

Therefore, it is hypothesized that:

Hypothesis 4a: The range of the receiver moderates the relationship between cohesion and receipt of knowledge transfer, in such a way that it decreases the optimal degree of network cohesion for the receipt of useful knowledge.

Hypothesis 4b: The range of the sender moderates the relationship between cohesion and receipt of knowledge transfer, such that it decreases the optimal degree of network cohesion for the receipt of useful knowledge.

2.5.4 Cognitive distance

Cognitive distance, a relational characteristic of dyadic knowledge flow, relates to the extent that people "interpret, understand and evaluate the world" (Nooteboom, Van Haverbeke, Duysters, Gilsing & Van den Oord, 2007, p. 1017) in a different way. Different backgrounds or environments in which they developed their cognitive frame of reference can influence the way people interact with new knowledge, and as a result, can impact the ability of two people to exchange knowledge effectively. As a result, literature has shown that a degree of shared knowledge is needed to share ideas, and common knowledge has shown to improve the transfer of knowledge (Hislop, 2009), and actors' abilities to assimilate knowledge (Cohen & Levinthal, 1990; Boschma, 2005). Boschma (2005) and Nooteboom et al. (2007) argued that there is a duality to cognitive distance, where commonality is needed for mutual understanding, but distance is needed for the development of distinctly different and novel perspectives. Since cognitive distance makes it harder to transfer and subsequently use knowledge, and cohesive structures are more effective at fine grained diffusion of knowledge (Coleman, 1988), high cognitive distance will increase the need of cohesion to bridge the knowledge gap (Tortoriello et al., 2012). To the contrary, two cognitively close individuals will be able to transfer knowledge more effectively, regardless of the network-structure in which they are embedded. This low level of friction decreases the ability of cohesion to alleviate friction, and increases the informational redundance two cognitively close actors already have. As such, it is therefore hypothesized that:

Hypothesis 5: Cognitive distance moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in cognitive distance increases the optimal degree of network cohesion for receipt of useful knowledge.

2.5.5 Complexity

Not all knowledge is created equal, and its level of complexity influences the degree two which dyads can effectively exchange the underlying ideas, especially for successful knowledge application. Correspondingly, complexity has been argued to increase friction in dyadic transfer (Ghosh & Rosenkopf, 2015). The reconstruction of knowledge that takes place when knowledge moves between people, in the interpretation and the recombination that the receiver has to accomplish in order to successfully use knowledge, becomes harder as separate components of that knowledge are more interdependent(Hansen, 1999) and ambiguous (Szulanski, 1996). Especially complex knowledge has to be reconstructed at every transfer step in the network, taking effort, time and inhibiting transfer (Rivkin, 2001; Sorenson et al., 2006). Zhao and Anand (2013) argue that cohesive structures are more effective at transferring knowledge of all complexities, but that advantage grows as complexity increases. To the contrary, boundary spanner structures, where less ties are involved in bridging an organizational boundary, incur lower collective costs in transferring knowledge across units. In a continuation of these propositions, Kim and Anand (2018) argued that with high complexity, decentralized, more cohesive network structures are less negatively impacted by replication difficulty due to knowledge complexity. Centralized network structures, with one boundary spanner, were better suited with low knowledge complexity (Kim & Anand, 2018).

Cohesive structures are expected to facilitate the transfer of complex knowledge better than sparse structures. Complexity is therefore expected to increase the need for cohesion for successful receipt of useful knowledge, and increase the optimal level of cohesion. It is expected that complexity maintains the inverted U-shape between cohesion and receipt of useful knowledge, however, as with very high complexity, costs and difficulty of transfer become too large to be facilitated by any kind of network closure (Sorenson et al., 2006). On the contrary, low levels of complexity will make replication and use of knowledge easier, reducing friction, and subsequently reducing the value of cohesion, lowering the optimal level of cohesion for the receipt of useful knowledge. Therefore, it is hypothesized that:

Hypothesis 6: Complexity moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in complexity increases the optimal degree of network cohesion for receipt of useful knowledge.

2.5.6 The theoretical model

This study sets out to test the inverted U-shape relationship between cohesion and the receipt of useful knowledge. It identified that the counterbalancing forces in cohesion are the removal of friction, and the introduction of redundance. It identified several key relational, structural, and knowledge characteristics, that impact the level of friction and redundance in dyadic knowledge application, and, as a result, are expected to impact the relationship of cohesion with the receipt of useful knowledge; they either shift the optimum to a lower level of cohesion, and a lower maximum of percieved usefulness of received knowledge, or they increase both, by increasing the need for cohesion. The resulting theoretical model is shown in figure 2.6.

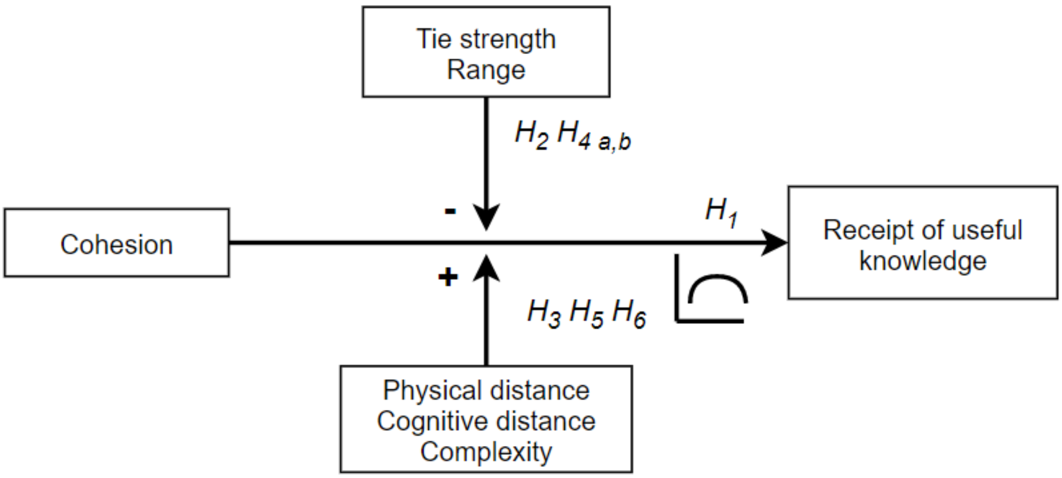


Figure 2.6: The theoretical model

This study set out to control for individual characteristics on an individual level, and dyadic level analysis of individual factors was done afterward, in a post-hoc analysis found in section 4.4.1. In the following chapters, this perspective will be outlined further, and will be motivated by introducing the perspective of friction and redundance in dyadic knowledge flow.

2.6 Academic relevance

In the pursuit of uncovering the relationship between cohesion and knowledge application, this study has academic relevance in several ways. The importance, and possible novelty of the model, are outlined below.

Curvilinearity

Curvilinear effects have been studied more in inter-organizational contexts, suggesting that the contextual factors play a bigger role between, rather then within organizations. While this seems to be the case (Wijk et al., 2008), the lack of such studies does not seem to be permissible. While curvelineair effects have been hypothesized and tested in other streams of related literature (Nooteboom et al., 2007), or proposed in simulation studies (Zhao & Anand, 2013; Kim & Anand, 2018), no studies of knowledge flow have tried to investigate the curvilinar relationship the balance between friction in redundance on the dyadic level yet.

Complexity

In a recent attempt to understand the organizational dynamics of complex knowledge flow, simulation studies have developed new theories on how complex knowledge diffuses within the organization (Zhao & Anand, 2013; Kim & Anand, 2018). In studies based on surveys of the perception of knowledge flow, the complexity of knowledge has not been studied recently. Seminal works in the field of dyadic knowledge flow (Reagans & McEvily, 2003; Levin & Cross, 2004; Tortoriello & Krackhardt, 2010) as well as more recent explorations (Levin et al., 2016; Nakauchi et al., 2017) have yet to incorporate knowledge complexity in their studies of the successful dissemination of organizational knowledge through interpersonal dyads, presenting an apparent gap in literature.

The extend to which simulated replication structures of knowledge resembles practice remains unknown, and no previous studies in the field of knowledge flow or transfer have developed a generalized item-set of survey questions that can be used to close the aforementioned gap in literature. To address this gap in literature, however, a couple of important questions need to be addressed. First, a thorough review of literature needs to confirm that, indeed, this measurement of dyadic knowledge complexity is not present in literature. Secondly, the subjectivist perspective on organizational knowledge (Nonaka & Peltokorpi, 2006; Hislop, 2009) suggests that it might not even be possible to develop a generalized item-set for the measurement of knowledge flow, and hence, the embedded and inherently indeterminate nature of knowledge in organizations (Tsoukas, 1996) might mean simulation and empirical measurement in survey studies never truly be reconcilable.

The usecase

To the best of the authors knowledge, this is the first study to investigate dyadic knowledge flow in an multi-unit NPD context, where all employees are part of a predefined interdependence of product knowledge, instead of through smaller project teams or consultancy activities. Social network characteristics of NPD teams have been investigated previously (Leenders & Dolfsma, 2016), and dyadic knowledge flow has been studied in a consultancy context (Reagans & McEvily, 2003), R&D organizations (Tortoriello et al., 2015), within company divisions (Levin & Cross, 2004), or across departments in a development company, but with a multitude of smaller projects (Nakauchi et al., 2017). No study has yet defined such a large and diverse predefined network of engineers, tied to the development of one product, in a network study before. Levin et al. (2016) tested a sample of multidisciplinary engineers working on one long-term project, but did not collect structural network data, and asked instead about the perception of closure, with a ego-network perspective. Smaller data splits are also frequent in literature, with Tasselli and Caimo (2019) evaluating two smaller networks with 29 and 21 participants, Tasselli et al. (2020) presenting a huge number of interactions originating from 40 managers spanning three measurement moments, and Levin and Cross (2004) finding their total 127 participants at three different companies.

Uncovering dyadic knowledge flow

This study aims to contribute to literature by providing a more nuanced view of dyadic knowledge flow. By specifically studying knowledge application, findings on transfer and acquisition are complimented. As (Phelps, Heidl & Wadhwa, 2012) put it, there is a need for studies to address how knowledge properties related to understanding an artifact influence adoption and diffusion on the interpersonal level. Furthermore, Phelps et al. (2012) calls for studies that “conduct more research examining the characteristics of knowledge being transferred, adopted, and created” (Phelps et al., 2012, p.1141). The present study aims to address these gaps in literature.

3. Methodology

This chapter describes the survey study conducted for this thesis. The chapter starts with a discussion on the design process of the survey in section 3.1. It then reflects on the data collection process in section 3.2, and elaborates on how variables and control variables were operationalized in section 3.3. Afterward, the resulting data is described and its validity evaluated. In section 3.4, likert-scale measures are validated with confirmatory factor analysis. Other validation metrics, such as missingness, outliers, multicollinearity, network validity and bias, are discussed in section 3.5. The chapter ends with section 3.6, a description of how the regression analyses were conducted.

3.1 Research design

The aim of this study is to investigate the how relational, structural and knowledge characteristics of a dyadic interaction impacts the flow of knowledge across the dyad, and enables the successful application of knowledge. A survey study was selected as the appropriate methodology to answer the research questions raised in this study. Network-data is usually gathered through surveys or through the evaluation of digital communication (Borgatti, Everett & Johnson, 2018). Since access to digital communication channels was not available in the case company, surveys were the only methodology that allowed for the elicitation of network data. Furthermore, the hypotheses surrounded the receipt of *useful* knowledge, it considered a subjective evaluation of the knowledge in its context, something best evaluated in a survey of the people who use the knowledge, and not found in secondary data available in the cross-departmental setting of the study.

The study concerns a study of cross-sectional data, since the data was gathered only once. Cross-sectional data was considered the best way to conduct network analysis in the use-case, given the effort required to get everyone aligned and participating in a single survey already. Response-rates across multiple surveys were deemed too much of a risk to consider for the scope of this study. The data was analyzed with multiple-regression analysis, a method suitable for measuring the strength of quantitative relationships between variables (Hair, Anderson, Babin & Black, 2010).

3.1.1 Sample definition

Due to the nature of the study, participants were pre-defined based on the following: their collaborative responsibility of the design, development, manufacturing and servicing of a system module of one of the products of ASML. The system module was selected together with major stakeholders within the business-line. The selected module was chosen because of its heterogeneous composition in terms of teams, locations and functional backgrounds. The bounded network also had a manageable size; enough to generate the data needed, but not too big to complicate appropriately bounding and surveying participants too much.

When the usecase was chosen, semi-structured interviews were conducted to verify and understand the interdependence between these groups, and to further specify the definition of all relevant subgroups. A picture of all subgroups was drawn up, and at every interview, this picture was shown, validated, and feedback was integrated into the picture for the next interview. Iteratively, a level of saturation in the network definition was reached this way, where groups and names of those who contributed to the system module were collected and retained, and those deemed not relevant dropped. Interviewees were additionally asked to provide the names of group leads within subgroups that would be able to provide a full list of participants. The network definition picture is included in appendix A, since it contains sensitive information.

In the third step, all group/team leads, roughly 15, were contacted to validate a list of participants relevant for them. This list was composed of the teams they were responsible for, or were assumed to know more about. This phase mostly lead to a confirmation of the boundaries sketched in the network definition during the interviews. The researcher, together with knowledge managers within ASML, decided on the exact network boundary, however, since interviewees and contacts did return some contradictory information. When knowledge managers agreed a point of saturation was reached, the final list was composed and send to the program manager of the business-line for final approval. This approval finalized the list of survey participants.

3.1.2 Sample iterations

Once the survey was sent out, some alterations were still made to the network definition. Two major stakeholders criticized the network definition, stating that the product-teams and the customer service groups were defined imperfectly, and that too many people were added too the sample. This was validated, corrected, and the participants in question were notified that their participation was no longer needed. None of the people mistakenly added had already participated in the survey.

During the data collection, managers were frequently contacted to discuss response rates. During these discussions, some further fine-tuning to the network definition was done. It always turned out that too many names were added, and names were only removed, or in a couple of instances altered, for example, when the person in question had left and been replaced. In a few cases, employees themselves indicated that they had nothing to do with the system module. They were not removed from the network definition, unless also verified by their manager that they should not be part of the sample. The bounded network pictures were updated as soon as new information became available on these alterations. A software group also had too many software-engineers initially included. Based on discussions with a production manager, another previously added group was dropped, because their manager did not deem them to be relevant to this study (and these arguments were validated, and seemed correct). Out of all employees removed during the study, only four had already participated, all from the production group. In total, the amount of participants was downsized from an initial definition of 322 to 236.

3.1.3 Design for common method bias

Important for the design of analysis methods is the the pre-assessment of, and preventive design for, method bias. Method bias refers to the situation where the way the study was designed influences results, instead of the variables that are being measured (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). During the development of the survey, several potentially problematic biases were identified. These possible issues, identified based on (Podsakoff et al., 2003) and (Borgatti et al., 2018), are shown in table 3.1.

Non-participation was a serious problem to be addressed in survey design, because it can lead to

Identified issues	Relates to	Mitigation strategy
Non-participation	Low response rate, omission of groups or key participants (Borgatti et al., 2018)	Managerial commitment and communication, elaborate sending strategy, pilot study
Difficulty of questions	Item demand characteristics (Podsakoff et al., 2003), retrospective errors (Borgatti et al., 2018)	Pilot study, question design ("take the time to recall..." questions, participation sessions
Measurement bias	Common-rater effects (Podsakoff et al., 2003)	Emphasis on identifying all interactions, pilot study, using validated measures from literature where possible
Rater-bias	Social desirability, leniency bias, predictor and criterion measured using the same medium, at the same time (Podsakoff et al., 2003), attribution errors (Borgatti et al., 2018)	Managerial commitment and communication, participation sessions, pilot study, privacy office and work-council alignment on privacy standards,clearly stating privacy considerations

Table 3.1: Survey design interventions to prevent method bias

the omission of cliques or central participants, which can have a large impact on findings (Borgatti, Carley & Krackhardt, 2006; Grannis, 2010; Borgatti et al., 2018). Major stakeholders were asked to continuously promote participation to ensure the required response rate was achieved. Since suitable data collection was crucial for this study, substantial effort was allocated to collection. A thorough description can be found in section 3.2.

The difficulty of questions was another potential issue, since network recall is notoriously difficult for people (Borgatti et al., 2018), and study design had to accompany good recall. More difficult survey questions, like these, can have influence on judgement of respondents (Podsakoff et al., 2003). Complexity item questions and question framing in general were rephrased multiple times during the pilot-study and in discussion with academic supervisors to improve comprehensibility.

Measurement biases, and rater bias, were the last area of concern in the design of the survey. Common method biases related to all items being measured in the same survey, were identified, but not completely addressed in this study, and remains a weakness of this study design. While some items might be subject to these biases, (Reagans & McEvily, 2003) argue that cohesion and range are not, because they are composite scores based on multiple raters. Still, especially receipt of knowledge, tacitness and complexity are subject to measurement and rater biases. Implications for the results and future research are discussed in section 5.2.4 and section 5.2.5.

3.1.4 Pilot study

To address a lot of the possible issues outlined above, a pilot study was conducted, to evaluate quality of survey questions and descriptions, and to ask about participation intention. First, two subject matter experts helped improve the phrasing of questions. In the pilot study, two other subject matter experts, three participants part of the defined network, and two not part of the network, answered questions about the survey through a short questionnaire or a meeting (whichever they preferred). All employees had not seen the survey before, and were not involved in development. The findings of this pilot study are summarized in table 3.2.

Concern/issue	**No. of times mentioned**
Privacy/data sensitivity	3/5 employees indicated participants might be hesitant to provide the data required based on privacy concerns
Length of the survey	3/5 employees and 1/2 experts indicated survey was long
Rephrasing/framing the survey	5/5 employees and 2/2 experts had advice on phrasing
Variables/what is being measured	1/2 experts added two control variables

Table 3.2: Pilot test results

The main output of the pilot study was re-phasing most parts of the survey. Difficulty of questions, called 'item demand characteristics' (Podsakoff et al., 2003), had the potential to significantly impact results. To adress this, the development of network questions is ideally done together with participants (Borgatti et al., 2018).

Due to the sensitivity of the information collected, and the impact that non-response could have had on findings, the privacy measures taken were communicated as clearly as possible in the communication to participants. The survey had been aligned with the privacy office and the privacy committee of the works-council beforehand, and this was actively communicated to participants.

3.1.5 Resulting survey

After several iterations, the survey structure was as shown in figure 3.1. The survey was compartmentalized in three parts, batching questions of a similar nature, aiding in comprehensibility. The survey was completely constructed using Microsoft Infopath, an older, but more customizable tool than the alternatives available, and embedded in Microsoft Sharepoint for participation and data storage. Data was downloaded in excel, and processed and analyzed using Python through JupyterLab, primarily using statistical analysis packages Statsmodels and PySci and network analysis packages NetworkX and Pyvis. Section 3.3 on measures further elaborates on the reasoning behind the phrasing from an academic standpoint, and appendix B provides an overview of the complete survey.

INFORMAL KNOWLEDGE SHARING NETWORK [SYSTEM MODULE]
Introductionary text, explaining the goals of the study
Part 1 – Tell us about you
Questions on gender, age, field of expertise, areas of expertise, educational degree, part of the organization, experience in current sector, part of product team, in managerial role, primary workspace building, which floor, experience in other sectors
Part 2 – Your personal network related to [system module]
Explanation about network questions. Request to list all interactions, also unsuccessful
Pictures with all names of the predefined network, sorted by sector, team and last name
List were people could enter all the names they had asked advice in the last three months. For every person, respondents were asked to indicate closeness and frequency of interaction
Part 3 – Evaluating interactions
Explanation about next segment. Emphasis on confidentiality
Four sections, one per interaction, two optional, in following order: One successful, one less successful, one successful, one less successful

Figure 3.1: Survey overview

3.2 Data collection

Data was collected over a period of two months. This included a main period of data collection, of 6 weeks and one day, in which all participants were asked to participate, and a much shorter, second round, in which ten influential participants, deemed important for the representation of the network, were asked again to participate.

During data collection, it became rapidly clear that reaching the intended 80% response-rate was going to be a formidable challenge. While the originally planned data collection period was three weeks, in the first two weeks, only 20% of the needed, and 16% of the total population of the bounded network, was collected. Respondents noted that time-pressure, non-prioritization by their direct supervisors and the significant length of the survey prevented them from participation. The latter stood out, as many respondents, when asked, indicated that they had opened or started the survey, but were reluctant to finish the survey that for many took 30 minutes or more to complete. In response, only two interaction evaluations, instead of four, were required after three weeks, to decrease the time investment required from participants.

Many different data collection strategies were applied in parallel to stimulate participation. These included information sessions, weekly planned Q&A participation sessions, where respondents could join, participate, and ask questions during the completion of the survey. These sessions were planned in all participants' calendars at differing times during the week, over a period of four weeks. For some individuals, individual survey participation sessions were planned. Stakeholders were contacted to discuss response rates, most stakeholders responded by sending a reminder email to the groups they were responsible for.

When all of these measures did not seem to work, a more personal approach was used. The office in Europe, where a lot of people where situated, was visited to discuss the study with

Method	Description
Major stakeholder reminder email	Three, one at send-out and two follow-up reminders from the project sponsor.
Participation session	Organized four sessions per week for four week, inviting all participants. 31 people participated in one of these sessions.
Group meeting presentation	Presented during two group meetings, one for all groups in Asia.
Reminders to group-leads	Continuous, team leads were contacted if RR was lower than average and aligned on how to increase response.
Office visit	Office was visited thrice to talk to participants.

Table 3.3: Data collection initiatives undertaken to improve the response rate

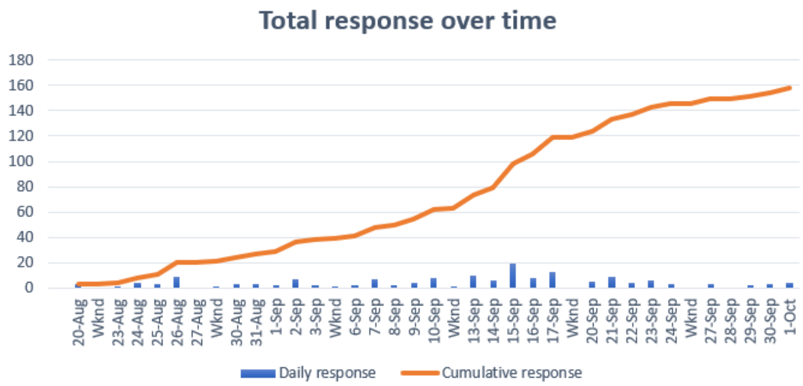


Figure 3.2: Survey response rate over time

participants and to see what would help them participate. A sheet was drawn up of all network members and the correspondence with said member (what they had said about participation). Based on this sheet, personal reminders were sent on Microsoft Teams to everyone who had not yet participated, prioritized by when they would go on holiday. This strategy was far more effective than stakeholder emails and sessions people could participate in out of their own initiative. Additionally, some group leads allowed a presentation during group meetings about the study. This also boosted responses in some groups significantly, especially in the participant group in Asia, where the meeting was interpreted as manager buy-in. A summary of the communication methods can be found in table 3.3. While it was not clear what the exact effect was, making two interaction evaluations optional was also deemed effective, as it reduced participation time significantly. All effort considered, data collection took considerably more time than expected. The data collection time-line, with the associated cumulative responses is, is shown in figure 3.2

After data collection ended, ten individuals were contacted one last time, based on how many people had evaluated interactions with them, to address as much missing data as possible. These requests yielded an additional 3 participants. After the addition of this data, a point of saturation was clearly reached, and data collection was finally closed.

3.3 Measures

This section discusses the operationalization of all variables used in the regression analysis and network data. First, the solicitation of network data is discussed. Afterward, the reasoning behind all other variables is provided. This section ends with a set of operationalization tables, where the measurement of each variable is specified further. This is table 3.4 for all variables in the hypotheses, and table 3.5 for all control variables.

3.3.1 Solicitation of network data

This study concerns a 'whole-network' survey study, where respondents can indicate which colleagues, of the predetermined set of alters described in the network boundary presented in section 3.1, they have asked advice or knowledge related to the system module. This method is chosen for two reasons. First, it will help with recall issues, as respondents are notoriously bad at recalling names of people at one given moment, even when they have worked with them closely for 8 months (Borgatti et al., 2018). Secondly, the aim of the study is to study a group of people interdependent on knowledge on the same system module; a bounded network of this group should naturally represent a solid selection of most of the important network ties (Hennig, Brandes, Pfeffer & Mergel, 2012).

Just like in the study of Reagans and McEvily (2003), however, respondents were given the opportunity to identify alters not on the primary list. This was done because network definition efforts has shown that the boundaries of such a multi-departmental NPD-network were quite fuzzy, and while snowballing, some artificial saturation point had to be chosen, consciously exposing the network definition to unavoidable imperfections. Hennig et al. (2012) show that, ideally, all actors have to be identified for a complete-network study. Stakeholders indicated that all groups were also somewhat interdependent on other groups outside of the bounded network. 'Extra' alters were not added to the dataset themselves, and did not count towards network characteristics like cohesion. These alters were a back-up, in case it was discovered that a serious mistake has been made in the bounding of the network, and a group had to be added. It was decided that this additional data would not be used for analysis, except for construct measurement validation.

Respondents were asked to specifically list advice-seeking interactions, in line with (Levin & Cross, 2004), on the system module as a subject representation of shared underlying knowledge. A critical assumption in this study is that this was the most appropriate elicitation of network ties over which important, dependence related knowledge would flow. "Advice/expertise" seeking interactions were chosen to elicit those interactions that concerned the exchange of knowledge needed for the work on the common interdependent knowledge asset. Subsequently, the network data was gathered by asking the following question:

"Which of the people below have you asked for their advice/expertise about the [system module], in the last three months?"

The question is accompanied with a couple of instructions, like:

"Regardless of whether these interactions have been successful or unsuccessful, we would like you to indicate all of them, no matter the results."

and

"Please take a couple of minutes to look through the two lists below and reflect. Below, you will be able to indicate any name you have not found in the list."

Respondents were provided a timeframe for in which they were asked to recall interactions. This timeframe was "in the last three months", in accordance with Levin and Cross (2004). This timeframe was selected, given the dynamic nature of the network, and because asking respondents to recall all relevant connections over a longer period might be too straining, leading to incompleteness issues. In contrary of Levin and Cross (2004), however, respondents were asked for all advice seeking relationships, and not only of one project, given the cross-departmental nature of this survey-study.

The size of the bounded network was considerable, with 236 names provided for participants to choose from. An overview of all names was provided to participants, categorized per department, and ordered alphabetically, to ease the search, which is important in network recall accuracy (Borgatti et al., 2018). Just like (Reagans & McEvily, 2003), a maximum of 20 alters was set, because otherwise the measurement of tie strength would have become very time-consuming for respondents.

3.3.2 Receipt of useful knowledge

Receipt of useful knowledge was operationalized with the use of the composite item set developed and validated by Levin and Cross (2004). The choice for this variable is motivated in section 2.2. Survey items include: information/advice I received from this person made (or is likely to make) the following contribution to (1) client satisfaction, (2) project team performance, (3) the project's value to my organization (4) this projects quality, (5) the project's coming in on budget or closer to coming in on budget (6) reducing costs in the project, (7) my ability to spend less time on the project, (8) shortening the time this project took. Items were measured on a 7-point likert scale ranging from "Contributed very negatively" to "Contributed very positively".

3.3.3 Cohesion

Cohesion was operationalized in accordance with to Reagans and McEvily (2003) and Tortoriello et al. (2012), with Burt's network constraint. Network constraint is among the most used constructs in network analysis literature, and as a result, deemed highly reliable. Cohesion measures the degree of triadic closure between a respondent and their alter, by summing over the product of tie strengths between the set of shared connections.

3.3.4 Tie strength

Tie strength is a broad concept, composing many attributes (Marsden & Campbell, 1984). Most studies, in accordance with the findings that closeness and frequency of interaction are the most important characteristics, use close variants of that definition (Hansen, 1999; Reagans & McEvily, 2003; Levin & Cross, 2004; Tortoriello et al., 2012). Tie strength can be evaluated from the source (Reagans & McEvily, 2003) and recipient (Levin & Cross, 2004), which uses a roughly equivalent measure, or by using an average of both evaluations (Tortoriello et al., 2012). For this study, since it concerns a advice seeking network from one of the two perspectives like Tortoriello et al. (2012), averaging over both perspectives is arguably more accurate, but not possible. Instead, the one-sided measure of (Reagans & McEvily, 2003) is adopted. It was deemed more straight forward and easier to interpret and implement in the survey than the measure used by Levin and Cross (2004). It had also been tested in combination with cohesion and range before, easing integration and increasing confidence in the validity of these measures. Values of frequency and closeness were summed and averaged, after which relative tie strength was computed with by dividing the reported tie-strength the total tie strength scores reported by the participants (Reagans & McEvily, 2003).

3.3.5 Physical distance

Based on the context of the study, there are different ways in which physical distance might interact with the receipt of useful knowledge. All methods found in literature, implicitly assume how distance should be modeled; either giving an ordinal scale to different configurations (Levin & Cross, 2004) or giving a linear attribution to distance (Kleinbaum et al., 2013). Since the network studied had a variety of locations and distances, it was assumed that geographic distance was the most nuanced way to represent distances between people. Both an ordinal scale and log-transformed distance in meters were tested, and in the end, the ordinal variable proved more suitable for the study, because of its increased predictive ability for the receipt of useful knowledge and the fact that log-transformed distances resulted in very messy data. Proximity was operationalized using the floor, building, and office site in ordinal intervals.

3.3.6 Cognitive distance

For cognitive distance, five separate measures were included in the survey. Four measures were taken from literature, and included a measure of educational degree similarity (Reagans & McEvily, 2003), difference in field of expertise (Reagans & McEvily, 2003), similarity in expertise areas (Reagans & McEvily, 2003) and the department of employment (context specific, but similar to Tortoriello et al. (2012)).

Another variable, departmental experience, was added, which was deemed a meaningful addition to departmental membership measure of Tortoriello et al. (2012), based on the context. In this measure respondents were asked to report previous experience in years in other departments. In total, five variables were included for cognitive distance, which it quite considerable. It was decided to include this many variables because some of the variables, mostly departmental experience and area of expertise, were very sensitive to the correct interpretation of the question. Unfortunately, the resulting data showed that the data collected for these two variables was indeed too inconsistent and incomprehensible to use for this study.

3.3.7 Complexity

Complexity has not been used to describe dyadic-level exchange of knowledge in a survey itemset before. It has been studied in surveys before, in the context of transfer. Hansen (1999) discusses complexity as dependent knowledge and asks whether the product or component worked on stands alone or is it dependent on other components or products. Zander and Kogut (1995) look at complexity in the multitude of manufacturing processes that are important for a specific process, implying that an increased amount of important areas models how interdependent these areas are. Complexity is also studied in patents, where it represents the number of subclasses of a specific patent (Sorenson et al., 2006; Ganco, 2013), the ease of replication of those sub-classes (Sorenson et al., 2006) and the uniqueness of the combination of those sub-classes versus their individual occurrence (Ganco, 2013). In simulation, it is represented by the size of the set of expertise areas, and the matrix of the interdependencies between those expertise areas (Zhao & Anand, 2013; Kim & Anand, 2018).

To incorporate these aspects of knowledge complexity in a survey item-set, these four aspects have been translated to the ASML context: the uniqueness/specificity of the knowledge transferred, the amount of knowledge needed to replicate it, the interdisciplinarity, or number of distinct subsets of knowledge, and the dependence different types of knowledge needed to replicate it. Therefore, the proposed survey questions are:

- You do not have to be a domain expert to be able to understand and apply this knowledge *(reverse scale, uniqueness)*
- This knowledge is specific for the domain in which I work *(uniqueness)*
- A lot of prerequisite knowledge, or a large list of instructions, is needed to effectively understand and apply this knowledge. *(amount of components)*
- This knowledge does not depend on a large amount of different areas of expertise to be understood and applied. *(reverse, diversity of components)*
- The work that I needed advice on requires a lot of time spent on considering how different aspects or components of the work interact and influence each-other. *(interdependence of components)*
- The work that I needed advice on is not highly dependent on different disciplines, interests or perspectives. *(reverse, interdependence of components)*

These questions were iterated upon together with university supervisors and ASML employees, and also specifically asked about in the pilot study, where employees deemed them somewhat difficult, but clear enough to answer. As a sanity check, and to further validate the quality of this novel measure, a qualitative question was added to each interaction evaluation: “please describe in one sentence what you sought advice on”. The interaction descriptions will be used to sample descriptions together with their complexity scores, so the validity of the measure can be assessed qualitatively to some degree, next to the usual quantitative checks.

3.3.8 Range

Range is based on the network measure of (Reagans & McEvily, 2003), with one important difference. Network diversity measures a respondents’ connection across separate and distinctive areas of knowledge. To delineate different knowledge categories, (Reagans & McEvily, 2003) use the area of expertise variable discussed in section 3.3.6. In the data collected with the survey, however, this variable produced data too noisy to neatly categories areas of expertise. Knowledge managers agreed that this was due to the lack of congruence in terminologies across departments in the sample. As a result, departments were characterised as destincted groups of knowledge, and range was defined as network diversity across the departments.

3.3.9 Control variables

In order to study relationships between dependent variables and independent variables, and to derive conclusions on the relationships between them, the inclusion of control variables is crucial. Control variables enable conclusions to be drawn on the effect of independent variables on dependent variables, holding control variables constant. Doing this increases confidence in the validity of conclusions drawn from regression analysis.

Tacitness

Hansen (1999) showed that non-codified knowledge inhibited knowledge transfer between organizational units. Hansen and Løvås (2004) argued the non-codification of knowledge contributes to its complexity. The present study uses a different definition of complexity, as outlined in section 3.3.7. Still, it is important to control for non-codified knowledge, since this can make transfer harder, and might have a similar, but separate, effect than the complexity item-set introduced in the present study.

Tenure

Multiple respondents have indicated that staying at ASML for a long time can have a significant impact on the strength and size of someone's personal network. As such, tenure was expected to relate to the receipt of useful knowledge. Many network studies control for tenure, recognizing the importance of tenure as a control variable (Levin & Cross, 2004; Tortoriello & Krackhardt, 2010; Tortoriello et al., 2012; Tasselli & Caimo, 2019). As a result, respondents were asked how many years they have worked at ASML.

Network size

Similarly, network size was deemed a possible predictor of the receipt of useful knowledge. Many network studies control for network size, recognizing its importance as a control variable (Tortoriello & Krackhardt, 2010; Tortoriello et al., 2012). Network size of respondents is operationalized as the number of ties indicated in the study by every given respondent.

Gender

The theory of homophily shows that similarity increases likeliness of communication frequency and transfer success (McPherson et al., 2001). This study considers a lot of characteristics of similarity (cognitive, geographical). As (Kleinbaum et al., 2013) have shown, however, gender is an additional important aspect of similarity that could explain the receipt of useful knowledge. Gender has been used as a control variable in comparable network studies (Levin & Cross, 2004; Tortoriello et al., 2012; Tasselli & Caimo, 2019).

Age

The same principles outlined for gender also count for age (McPherson et al., 2001). While age is less often connected to the receipt of useful knowledge, it is still often used as a control variable in comparable network studies (Reagans & McEvily, 2003; Levin & Cross, 2004; Tasselli & Caimo, 2019). As a result, respondents were asked to report their age in years.

Friendship

Reagans and McEvily (2003) showed how friendship can be a significant predictor of successful transfer of knowledge, separate from comparable variables like tie strength, making it an important control variable. As such, a question was added on whether people like to spend time together outside of work, for each evaluated interaction.

Managerial role
Based on a conversation with a knowledge manager, two other controls were added. The first one was managerial role, or whether someone had a role of either a group lead or a project manager. The knowledge manager expected that managers might have a different frame of reference, especially given the fact that the study concerns organizational boundaries, and it was reasoned that managers might have more boundary-spanning functions than non-managers.

Product-team membership
The second control variable concerned product team membership. Product teams are cross-departmental teams, and are the teams that make decisions in the NPD process, and have the responsibility to combine and integrate all relevant perspectives for these decisions. As a result, product team membership was deemed relevant, and added to the study.

Unfortunately, based on contextual expertise, the data resulting from this measure was deemed incorrect. A lot of employees interpreted the term ‘product team’ to mean something else then what it formally represents within the company, and many more ‘yes’-answers were reported then should have been possible. Subsequently, this variable was removed from analysis.

3.3.10 Operationalization tables

Variable	Definition	Measures	Rating
Receipt of useful knowledge	Adopted from (Levin & Cross, 2004)	Information/advice I received from this person made (or is likely to make) the following contribution to: (1) client satisfaction (2) project team performance (3) the project's value to my organization (4) this projects quality (5) the project's coming in on budge or closer to coming in on budget (6) reducing costs in the project (7) my ability to spend less time on the project (8) shortening the time this project took	7-point likert scale ranging from "Contributed very negatively" to "Contributed very positively"
Geographical proximity	Inspired by (Levin & Cross, 2004)	(1) Which ASML building is your primary workspace? (2) Which floor?	(1) Respondents could select a predefined building from a list, or provide one themselves. (2) Respondents could provide an integer
Cognitive distance	Adopted from (Reagans & McEvily, 2003) and the research context.	Question on functional background: (1) What describes your field of expertise best? (feel free to add a different top level discipline) Question on part of organization: (2) In which part of the ASML organization do you work currently?	(1) Respondents could select a predefined field of expertise from a list, or provide one themselves. (2) Respondents could select a predefined department from a list, or provide one themselves.
Complexity	Self-developed, compiled from (Zander & Kogut, 1995; Hansen, 1999; Sorenson et al., 2006; Zhao & Anand, 2013)	(1) You do not have to be a domain expert to be able to understand and apply this knowledge - *reverse scale* (2) A lot of prerequisite knowledge, or a large list of instructions, is needed to effectively understand and apply this knowledge (3) This knowledge is specific for the domain in which I work (4) The work that I needed advice on requires a lot of time spent on considering how different aspects or components of the work interaction and influence each-other (5) The work I needed advice on is not highly dependent on different disciplines, interests or perspectives - *reverse scale* (6) This knowledge does not depend on a large amount of different areas of expertise to be understood and applied - *reverse scale*	"Indicate to what extent you agree with the following statements about the knowledge you sought from this person" - 5-point likert scale ranging from "Agree" to "Disagree"

Variable	Definition	Measures	Rating
Tie-strength	Adopted from (Reagans & McEvily, 2003)	Questions per identified contact: (1) How often do you interact with this person? (2) How close do you feel to this person? Tie-strength calculation based on tie-intensity scores: $p_{ij} = \frac{z_{ij}}{\sum_{q=1}^{N} z_{iq,q=j}}$ Where: p_{ij} is tie strength z_{ij} is tie intensity betwen ego and alter in dyad z_{iq} is tie intensity between ego and all alters ego is connected to	(1) 4-point likert scale, including "Daily", "Weekly", "Monthly" and "Less often/did not interact before". (2) 4-point likert scale, including "Especially close", "Close", "Less than close" and "Distant"
Cohesion	Adopted from (Reagans & McEvily, 2003)	Cohesion calculation based on tie-strength scores: $c_{ij} = \sum_{q=1}^{N} p_{iq} p_{qj,q \neq j}$ Where: c_{ij} is cohesion p_{iq} is tie intensity between ego and the triadic contact p_{qj} is tie intensity between alter and the triadic contact	Based on tie-strength values

Variable	Definition	Measures
Range	Network diversity. Adopted from (Reagans & McEvily, 2003)	Range calculation based on tie-intensity values and department-membership. Made up of two components, the strength of a persons' connection with a department, and interdepartmental connection. Network diversity: $nd_i = \sum_{k=1}^{6} p_k p_{ik}^2$ Where: nd_i is network diversity of person i p_k is interdepartmental connection, the strength of departmental connection within department k, p_{ik} is the strength of connection between person i and department k Interdepartmental connection: $p_k = \frac{\sum_{j=1}^{M_k} z_{ij}}{\sum_{q=1}^{S_k} z_{iq,q=j}}$ Where: S_k is the number of interactions identified by respondents in department k M_k is the number of people in department k z_{iq} is tie intensity between an individual within department k and an alter z_{ij} is tie intensity between an individual within department k and an alter Strength between person and department: $p_{ik} = \frac{\sum_{j=1}^{Nk} z_{ij}}{\sum_{q=1}^{N} z_{iq,q=j}}$ Where: Nk is the number of contacts of person i in department K N is the number of contacts of person i z_{iq} is tie intensity between person i with alter q z_{ij} is tie intensity between person i with alter j in department k

Table 3.4: The operationalization of variables

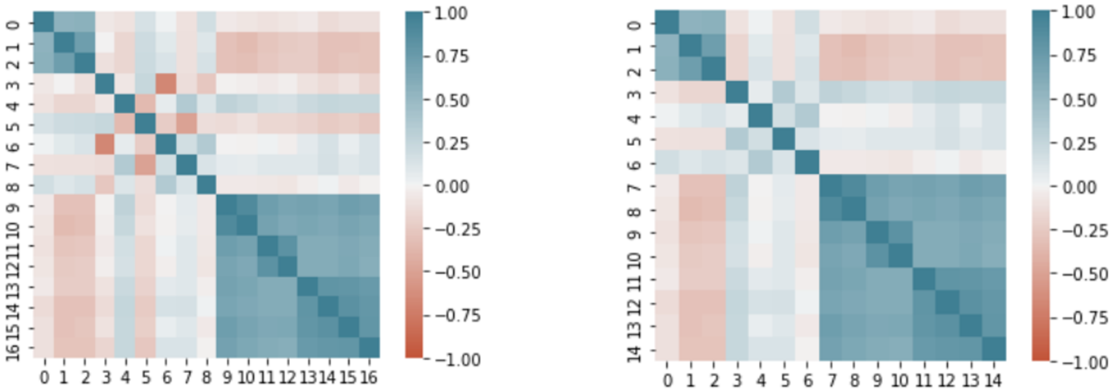
Control variable	Definition	Measures	Rating
Tacitness	Adopted from (Hansen, 1999; Levin & Cross, 2004)	(1) How well documented was the knowledge you received from this person? (2) Was all this knowledge sufficiently explained to you in writing? (3) What type of knowledge did you receive from this person?	Three 7-point likert scales. (1) It was very well documented ... It was not well documented (2) All of it was ... none of it was (3) Mainly reports, documents, manuals ... mainly personal practical know-how
Gender	The gender of the respondent	What is your gender?	Female = 0 Male = 1 Other/wish not to disclose = 2
Age	The age of the respondent	What is your age?	Respondents could provide an integer
Tenure	The tenure of the respondent	Question tenure current department: (1) How many years of experience do you have working in that department (round to nearest integer)? Question on experience in other departments: (2) Please indicate below, for each department in which you also have working experience, the amount of years you have worked there.	(1) Respondents could provide an integer (2) Respondents could enter eachdepartment and an integer in a list
Managerial role	Whether the respondent is in a manegerial role, either as a project lead, or as a group-lead	Are you currently in a managerial role (GL, PO or other)?	No = 0 Yes = 1
Product team membership	Whether the respondent is part of the product team, the team that makes go-no go decisions on the NPD stage-gate process	Are you currently part of a product team?	No = 0 Yes = 1
Networksize	The total amount of connections of a respondent, both incoming (reported by respondent) and outgoing (reported by other respondents)	Summation of ties in list of neighbours in the network	Based on network data

Table 3.5: The operationalization of control variables

3.4 Factor analysis

Confirmatory factor analysis (CFA) was used to determine whether the measured items of the three Likert-scale variables, receipt of useful knowledge, complexity and tacitness, measure the intended latent construct. For this validation, the entire dataset of 423 observations is used, so also observations that were not suitable for regression because they were missing data. A correlation matrix between items, showed in revealed that, while tacitness and the receipt of useful knowledge, as anticipated, are highly inter-correlated, the observed variables of the complexity are not. Two specific items, one corrected from a reverse scale and one regular variable, which were both designed to measure interdependence of knowledge, showed to be especially uncorrelated with other items. In factor analysis with Varimax rotation, these two items produced negative factor-loadings on all three factors, highlighting severe differences in measured latent constructs then all other questions.

The removal of the two interdependence items, which were proving unsuitable for the measurement of a homogeneous complexity construct, improved overall performance. The remaining four items had low but present inter-correlations rating from 0.05 to 0.34. The difference is shown in figure 3.3. In Factor analysis with Varimax rotation showed sub-par loadings on the third factor, ranging between 0.323 and 0.513, as seen in table 3.6. All factor loadings would preferably lie above 0.5 for suitable internal consistency (Hair et al., 2010).



(a) Correlation matrix before removal two complexity items

(b) Correlation matrix after removal two complexity items

Figure 3.3: Correlation matrices of individual likert-scale items

Based on these factor loadings, Cronbach's alpha, Average Variance Extracted (AVE) and Construct Reliability (CR) were calculated for all three latent constructs. These are shown in table 3.7, with their indicated rules of thumb by Hair et al. (2010). While rules of thumb are hard cut-off scores and absolute proof of reliability, the combination of the good results in the reliability tests, and the established and tested nature of Tacitness and the Receipt of useful knowledge, provides confidence in these measures' reliability and validity. Given the factor loadings and low reliability scores, complexity has to be rejected as a suitable measure for regression analysis. Whether or not the construct has been designed to measure the right construct (which is a more qualitative discussion), its internal validity is too feeble to use it for regression analysis. Hypotheses concerning complexity hence cannot be tested.

	VARIMAX-ROTATED LOADINGS		
	Factors		
Variables	**1**	**2**	**3**
Type of knowledge		*0.673*	
Explained in writing		*0.793*	
Well documented		*0.827*	
Uniqueness			*0.323*
Heterogeneity			*0.513*
Preliminary knowledge			*0.401*
Uniqueness			*0.459*
Shortning time	*0.868*		
Less time	*0.822*		
Reducing	*0.797*		
Budget	*0.781*		
Quality	*0.824*		
Organizational value	*0.816*		
Team performance	*0.862*		
Client satisfaction	*0.809*		

Table 3.6: Factor loadings of the individual likert-scale items
Factor loadings less than .30 are not shown, the three largest values not shown are .268, .216 and .156

	α	*conf. int.*	**AVE**	**CR**
Tacitness	0.82*	*0.79 - 0.85**	0.59*	0.7*
Complexity	0.55	*0.48 - 0.61*	0.18	0.63
Receipt of usefull knowledge	0.93*	*0.92 - 0.94**	0.68*	0.86*
Rule of thumb	0.7		0.5	0.7

Table 3.7: Reliability analysis factors

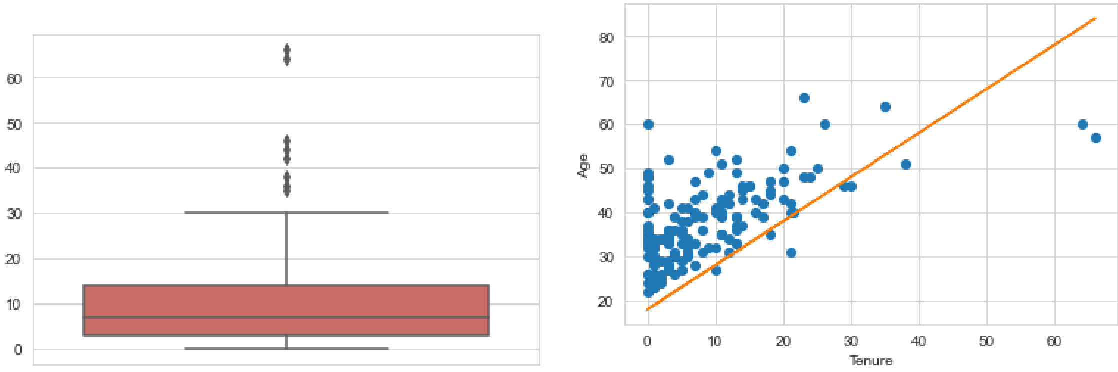
3.5 Data validation

Once data was integrated in variables, several data validation checks were preformed to evaluate how suitable the data was for regression analysis. The results of these checks are presented in this section.

3.5.1 Outliers

Three variables had outliers that needed to be addressed. The tenure and age variables had univariate outliers; both had contextually impossible values. Four respondents had indicated an age of 0, and one had indicated an age of 100. These values were resported as missing.

Tenure had six values above 60, while the age variable, after corrections mentioned above, ranged to 66. The tenure-variable had inherent problems in its integration in the survey; instead of one question asking total tenure, it was split into two questions. One question asked the tenure of the current department, the second other experience with ASML in other departments, in free form. This was done in order to measure the 'departmental experience' variable for cognitive distance. It seems that quite some participants misunderstood these questions, and duplicated (part of their) reported tenure. In 15 cases, there was enough evidence that participants had exactly doubled their tenure. In these cases, their tenure values were halved, and this was deemed a suitable correction for the problem in measurement. With this adjustment, however, two values above 60 remained, and it is possible that more incorrect tenure values remain in the data. Of the 'impossible' tenure values, the ones above 60 were removed from analysis. Other values right of the orange line in figure 3.4b were also judged to be invalid, but those were not removed. It was clear that tenure had noise in the data, and by removing only part of that noise, it would bias the data in a way that cannot be accounted for. Constant noise was deemed more desirable than very biased noise.

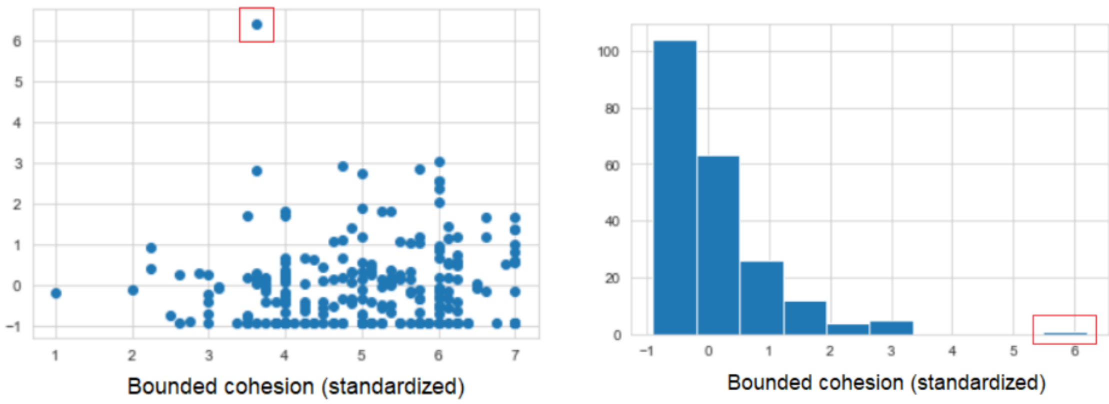


(a) Boxplot of tenure

(b) Scatterplot of age and tenure, with a line representing the tenure of someone starting at age 18. Values right of the line should not occur

Figure 3.4: Tenure data

All variables were screened for bi-variate outliers with a scatter matrix. This matrix only revealed one clear visual outlier, for cohesion, as shown in figure 3.5. This value, after being standardized, had a standard deviation of 6, which rendered its appearance hugely unlikely. Looking at what the data represented, however, it turned out that it was an expected result a dyad between two respondents of which 2/3rd of remaining alters were shared triads for the first respondent, and 2/5th of the remaining alters were shared triads for the second respondent. It is surprising that there is only one such case of high cohesion within the network. As a result, this outlier was initially not removed, since it is based on data from multiple respondents, and did not seem to be an extreme occurrence that cannot be explained.

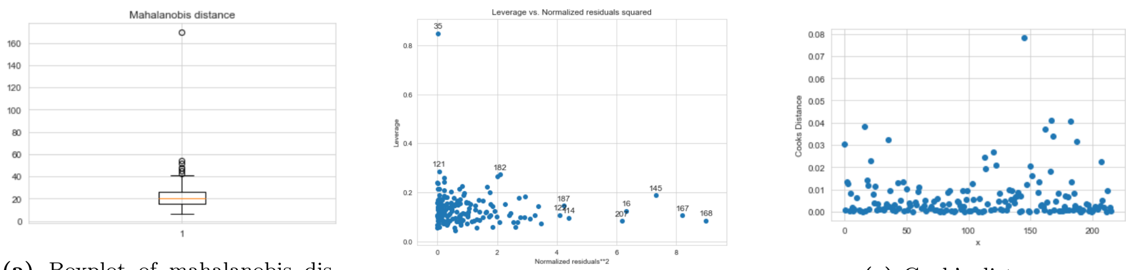


(a) Scatterplot cohesion against the receipt of useful knowledge

(b) Histogram standardized cohesion values

Figure 3.5: Multivariate outlier cohesion

To test for multivariate outliers, Mahalanobis distance was used, on all data used in regression analysis. In figure 3.6a a boxplot is shown of Mahalanobis distances. Based on (Hair et al., 2010) determination of outliers of $D^2/df > 4$, and df of 22 and outlier is classified as $D^2 > 88$. The figures show that the bi-variate outlier identified earlier and shown in figure 3.5, that seemed innocent and explainable, has significant impact on the estimation of model coefficients. This one data-point led to the significance of a variable that was completely insignificant after its removal. The data-point was deemed too influential, and therefore removed from analysis.



(a) Boxplot of mahalanobis distance

(b) Leverage plot

(c) Cook's distance

Figure 3.6: Multivariate outlier analysis

3.5.2 Missingness

There was a substantial amount of missingness present in the dataset, despite effort to prevent so by making fields in the survey required. All missing values are shown in table 3.8 below. Missingness in age was caused by four respondents not indicating their age, and three respondents not indicating their gender, as discussed in section 3.5.1.

Variable	Missing values
Tacitness	1
Age ego	9
Gender ego	6
Cohesion	37
Gender Alter	5
Tie-strength	13
Proximity	2
Cognitive distance	2
Total observations	239

Table 3.8: Missingness before processing

There were two important reasons why the missingness, for tie strength and cohesion was especially problematic, and had to be addressed. Firstly, there was a significant amount of missingness in the dataset in the complete interaction dataset with the limited sample of 239 observations. As Hair, Black, Babin and Anderson (2019) state, if the percentage of missingness surpasses the general rule of 10% missingness, it cannot be left unaddressed.

Secondly, a part of the missingness of the data was *not random*. During the survey study, a survey design flaw was uncovered that impacted missingness in a non-random way. In part 2 of the survey[1], participants were asked to indicate their knowledge seeking network, identifying contacts and the tie strength with that contact. In part 3, the participants were asked to evaluate interactions with some people from whom they had these advice-seeking interactions. These two data

[1]Recall that figure 3.1 provided an overview of the survey and its parts, for reference

entries were implicitly, and when found-out, explicitly connected, but not as such implemented, due to technical limitations of the survey tool used. It was possible for respondents to evaluate interactions with people whom they did not indicate were part of their knowledge seeking network, and [...] times a respondent made the mistake of not reporting a contact in their knowledge seeking network, omitting a tie strength evaluation. This missingness is non-random; the missing value could not be 0. Hence, this missingness had to be addressed.

The missingness in tie-strength was addressed with mean imputation. The average tie strength score that respondents reported for complete interactions, 1.70, was assigned to all values missing, only if they were named in the four interactions as a contact, but missed a tie strength evaluation in the identification of network contacts. New tie strength values were computed for all contacts of the ego of the missing tie, since the way tie strength is operationalized divides all reported tie strength scores by the total tie strength score for that respondent.

A relatively smaller amount of missingness in the reported tie strength values came from missingness in either the frequency or closeness missing. In the few cases where only one of the two were missing, the other value was used as a proxy of tie strength. Seven closeness and six frequency values were approximated by their corresponding counterpart. Combining these two methods removed all missingness in the data for cohesion and tie strength.

For tacitness, there was one missing data point. This point has been replaced by the average of the other two values. For one interaction evaluation no alter was indicated, leading to the removal of that evaluation as well.

The missingness in Proximity and cognitive distance was caused by a respondent selecting "other" for those categories, but not indicating the corresponding office or department. Rows with remaining missing data were removed from the data before regression analysis. The resulting missingness is shown in table 3.9.

Variable	**Missing values**
Age ego	9
Gender ego	6
Gender alter	5
Proximity	2
Cognitive distance	2
Observations missing a value	19
Total observations	220

Table 3.9: Missingness after processing

3.5.3 Assessing common method bias

As outlined in section 3.1.3, survey design exposed the measurement of constructs to method bias. This section investigates evidence of method bias in data distributions.

Measurement context effects

Measurement context effects, as named by (Podsakoff et al., 2003), refers to common method bias that can occur when variables are measured at the same time, through the same medium.

Complexity, tie strength, the friends dummy, tacitness, receipt of useful knowledge are all likert-scale evaluations in the same medium, at the same time.

As the regression results will show, later in this report, Tacitness, the only variable completely dependent on the likert-scale evaluation simultaneous to the evaluation of the dependent variable, is the most significant and robust predictor of the receipt of useful knowledge. This suggests strong prediction of the dependent variable (DV), but also raises red flags for common-rater effects, since those should be strongest for this variable as well. Figure 3.3 shows the inter-correlation between items of both likert-variables is consistent and negative.

One general way to investigate the presence of common method bias between constructs in a study is by the evaluation of confirmatory factor analysis (Podsakoff et al., 2003). Podsakoff et al. (2003) argue that in case all items load on a separate, common factor, this is evidence of strong bias in construct measurement. Luckily, this does not result from CFA, as all items weigh heavily only on their own expected factor, and there is no factor all items separately weigh on.

Item demand characteristics

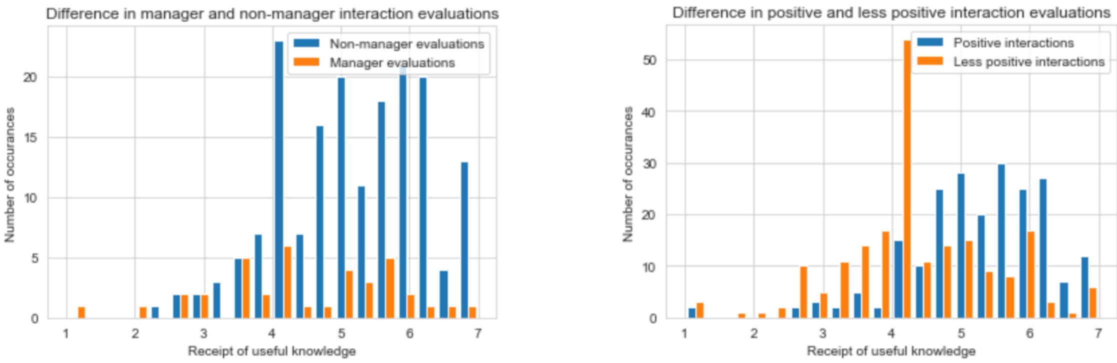
The likert-scale variables, receipt of useful knowledge, complexity, and tacitness, were suspect of item difficulty, another factor that can introduce common method bias (Podsakoff et al., 2003). One way this issue was dealt with was an investigation into straight-lining, which is viewed as a response to questions that are deemed too difficult. Observations were removed if all items were scored the middle score in the likert scale, or, in the case of the receipt of useful knowledge, the maximum score as well. For straight-lining, seven items were removed.

Social desirability

Social desirability seems to be a potential issue to this study. Discussions with participants during participation sessions showed that strong privacy guidelines, clear elaboration that results would not be reported on a personal level, and specific inquiry about "less-successful" interactions, did not prevent some respondents voice concerns that showed that some respondents viewed participation as an evaluation of colleagues. The resulting data (mean 5.05, std. 1.16 and skew -0.43) is averagely lower and less skewed, however, than that of Levin and Cross (2004) (mean 5.29, std. 1.09 and skew 0.75). Unlike Levin and Cross (2004), however, this study cannot introduce a fixed effect without over-fitting. Some tests were conducted to evaluate the impact of social desirability bias.

One striking finding was that managerial evaluations of the receipt of useful knowledge were more negative, and less skewed, than non-managerial evaluations, as shown in figure 3.7a. With an average score of 4.46 and skew of -0.41 for managers and average score of 5.18 and skew of -0.28 for non managers, the difference is not that big, but the amount of interactions evaluated by managers is higher than that of non-managers, while managers represent 15.3% of the sample.

Another finding was that "less positive" interactions, as they were phrased in the survey, suggested the prevalence of central tendency, as shown in figure 3.7b. This could be an indicator of social desirability bias, where in this case, respondents 'avoid the question'. Additionally, it is remarkable that quite some "less positive" scores were given especially high scores for the receipt



(a) Manager and non-manager evaluation **(b)** Positive vs less positive interaction evaluation

Figure 3.7: Negative skew (predominantly positive interaction evaluations) in the receipt of useful knowledge

of useful knowledge, another indication that the suggestion to rate more negative was ignored or avoided.

3.5.4 Non-response bias and network validity

Network validity can be evaluated based on response rates (Borgatti et al., 2018). As networks are highly sensitive to the effects of missing edges (Grannis, 2010; Borgatti et al., 2018), much higher response rates than ordinary regression studies are needed to accurately represent the sample. The response rate of this survey is 165 responses out of 236 employees in the bounded network, bringing the response-rate to 69,5%. Because the bounded network consists of many smaller teams of participants, the response-rates across teams are also deemed to be important to the networks' validity. Response rates of the groups, anonymized with a number, are shown in figure 3.8.

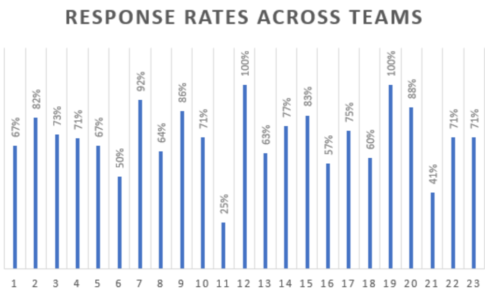


Figure 3.8: Response rates across teams

However, response rates do not say everything. Borgatti et al. (2006) show that the validity of networks' response rates depends highly on who was omitted, and the omission of top-respondents is much more impactful than people with a small role in the network. To investigate, given the context of knowledge seeking networks, it is assumed that the in-degree of nodes, the degree to which people form prominent knowledge sources as identified by respondents, are themselves part of the network. Based on this observation, two aspects of the in-degree of the network data are investigated.

Firstly, it was investigated whether the out-degree of respondents is significantly different from the out-degree of non-respondents. Since all network observations *are* independent, any deviations from the out-degree distribution is a sign that the data does not represent the bounded network. A Kolgomorov-Smirnov test was conducted, shown in table 3.10, and the null-hypothesis that both that respondents and non respondents have different out-degree values was rejected.

K-test statistic	P-value
0.09	0.81

Table 3.10: The Kolgomorov-Smirnov test of similarity in distribution of in-degree values for respondents and non-respondents

Secondly, the list of the top-contributors to the knowledge-seeking network, those with the highest out-degree values, was consulted. Of this list, 20% of the top-17%-contributors (all those with out-degree 10 or higher) were non-respondents. This suggested that there were no concerns for a significantly higher number of non-respondents in the set of top-contributors, for instance due to time-pressure or other priorities for these particular employees. The distribution of out-degree values is shown in figure 3.9.

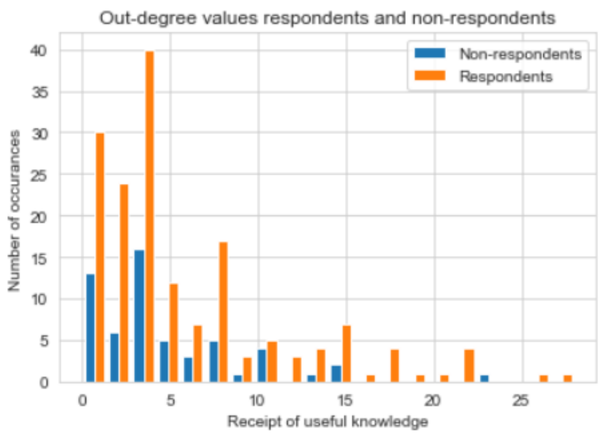


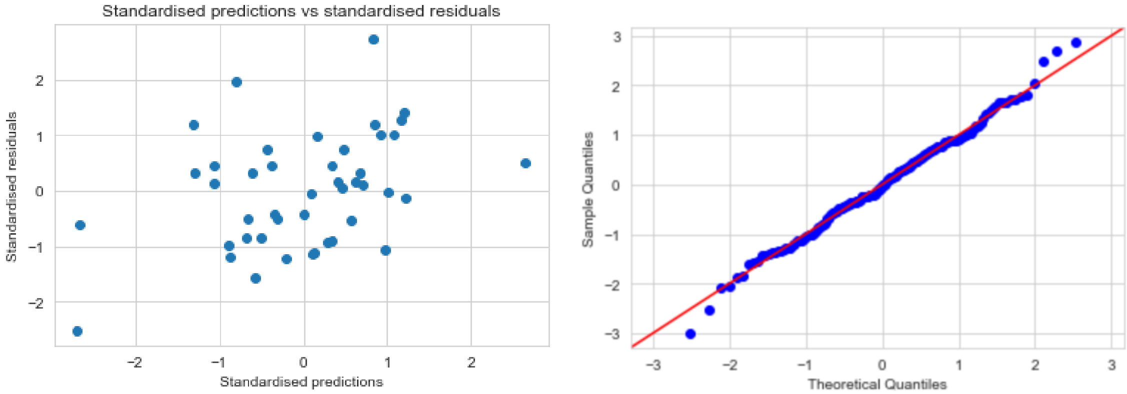
Figure 3.9: Histogram in-degree values respondents and non-respondents

3.6 Regression analysis

With data integrated and evaluated, regression analysis was performed. This section discusses the assumptions of regression analysis that need to hold for regression to be a suitable analysis method, elaborates on how models were tested, and discusses robustness.

3.6.1 The assumptions of regression

In order to be able to use linear regression for the testing of hypotheses, the validity of this analysis methodology as an appropriate method has to be checked. Multivariate regression analysis, which is to be used in this study, has several assumptions that the data must adhere too before the method can be deemed appropriate. These assumptions are the following: linearity, exogeneity, interdependence of observations, hetereoscedasticity and normally distributed error terms (Hair et al., 2010). Each of these assumptions is discussed in this section. To test these assumptions, the plots of residuals against predicted values, and a QQ-plot of the residuals, are used. These are shown below, in figure 3.10.



(a) Standardized residuals plotted against standardized predictions for a random five-fold split of the data

(b) QQ-plot of residuals

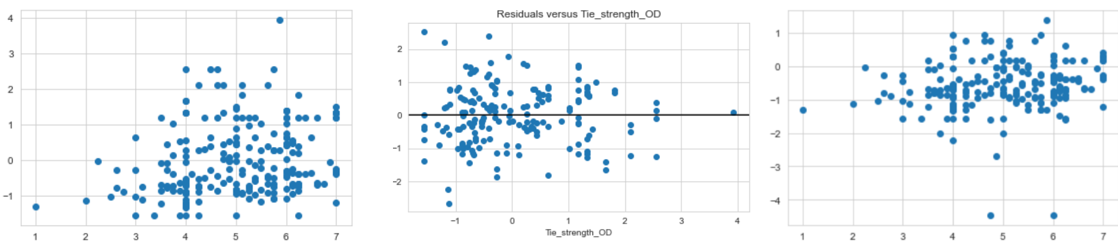
Figure 3.10: Residuals

Linearity

Regression analysis can only be done on relationships between dependent and independent variables that have a linear relationship, because non-linear relationships result in unreliable coefficient and error terms (Hair et al., 2010). Scatter plots, while not showing very strong linear relationships across the board and no clear patterns for some data with the receipt of useful knowledge, there are no non-linear effects visible, except for Cohesion, which is anticipated, and a quadratic term is introduced. The residuals, as shown in figure 3.10, do not show any signs of strong non-linearity.

Constant variance

The residuals plot does not show signs of inconstant variance, but is inconclusive. Additionally, the scatterplot for Tie-strength, and its partial residuals plot (as seen in figure 3.11 below), give concern for heteroskedasticity in the model due to the inclusion of this variable. Since the variable was found to be the significant predictor, where the log-transform was not, and the overall model does not show significant heteroskedasticity (see table 3.11), tie-strength was left in the model.



(a) Scatter-plot standardized Tie Strength (x-axis) and RUK (y-axis)

(b) Partial residuals plot for standardized tie-strength

(c) Tie-strength after log transformation

Figure 3.11: heteroskedasticity in tie-strength

To exclude issues due to heteroskedasticity in the overall model, a Breusch-Pegan test is conducted, which tests the null-hypothesis that there no heteroskedasticity in the regression model. Results show below indicate that the null-hypothesis could not be rejected, and hence, there is no evidence of heteroskedasticity, and tie strength did not need to be transformed.

Lagrange multiplier statistic	P-value
17.18	0.37

Table 3.11: Breush-Pagan test of heteroskedasticity

Independence of observations
Observations in the dataset were not independent. As respondents were able to evaluate multiple interactions, correlations between observations are expected for observations with the same rater. Non-interdependence between observations is a severe violation of the assumptions of simple least squares regression, and there is no simple remedy (Hair et al., 2010). Introducing a fixed effect would also have addressed this issue (Reagans & McEvily, 2003; Levin & Cross, 2004). Similarly, limiting every respondent to one interaction is also not an option, since it would lead to a very small dataset (with 165 observations) that would be subject to selection bias. Corrective interventions, such as introducing a variable that would compute the distance from average rating for a given respondent to model rater bias, introduce bias of their own. As a remedy, weighted least squares is used, where observations are weighted according to the amount of observations of a given respondent. This is discussed in 4.6.2.

Normality or errors
To test the normality of the error terms, a QQ-plot of the residuals is consulted, which is shown in figure 3.10b. This residuals plot shows that sample quantiles follow theoretical quantiles neatly, and hence, the normality of the error terms is confirmed.

No problematic multicollinearity
Collinearity, correlation between between the independent variables, is a potential problem for regression analysis, since it convolutes the interpretation of regression estimates. No evidence of problematic collinearity between independent variables was found in the correlation matrix. Multicollinearity, collinearity in a group of variables, was also ruled out, by the means of calculating VIF factors. The results are shown in table 3.12 below, and indicate that there is no evidence of problematic multicollinearity. Other models tested did also have no VIF values above 3.

Exogeneity
One of the assumptions of OLS is that variables are exogenous; they are uncorrelated with the error term (Hair et al., 2019). If they are correlated with the error term, OLS is no longer usable to estimate the β of the independent variable, because OLS will attribute some correlation between the error term and the independent variable as correlation between the independent variable and the dependent variable. Endogeneity would result in valid β estimates for prediction in the sample, but not for deducing causality. As a result, generalizability to other contexts would be confined. Levin and Cross (2004) and Reagans and McEvily (2003) solve this by introducing a fixed effect; a dummy variable for each of the respondents and contacts (which are also mostly respondents themselves). Any specific method bias, due to several variables coming from the same respondent, is also addressed, because this can be corrected by this fixed effect. Due to its limited sample size, this study does not have the luxury to introduce 165 respondent dummy variables, however.

Correlation between the independent variables and the error terms is not easy to test however, since, because of the possibility of endogenous variables, correlations between independent

Variable	VIF
Age ego	1.743842
Gender ego	1.099789
Friends dummy	1.109912
Tenure ego	1.655279
Age alter	2.045911
Gender alter	1.216584
Networksize ego	1.799053
Networksize alter	1.202431
Tenure alter	1.976989
Manager ego	1.377921
Manager alter	1.355915
Tie strength	1.696913
Physical distance	1.414578
Cohesion	1.381870
Same department	1.242368
Background similarity	1.286525
Educational dissimilarity	1.084731
Expertise overlap	1.318750
Complexity	1.161959
Tacitness	1.170753
constant	106.438373

Table 3.12: Variance inflation factors of the regression variables

variables and error terms are themselves biased. Two-stage least squares would be the most appropriate and robust method to adress endogeneity. Unfortunately, no suitable fixed effect was found to apply this method. Robustness checks introduced later in section 4.3.3 discuss how some models tested control for endogeneity to some extent. This study is still subject to limitations in the degree to which causal conclusions can be drawn from coefficient estimates(Hair et al., 2010). The implications of this are further discussed in section 5.2.4.

3.6.2 Model testing

Testing hypotheses

Although all variables have been carefully selected, simultaneous testing of variables runs the risk of errors in variable omission or inclusion (Hair et al., 2019). Therefore the hypotheses are tested through hierarchically built models. Example effects to test for are partial moderation and linearity/high order polynomial relations than quadratic. Below, regression equations are shown that are used in the control of these effects. These equations will be built up step-wise after all control variables are added to the model.

Testing for curvelineair effects

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon_1$$

$$Y = \beta_0 + \beta_2 X_1 + \beta_3 X_1^2 + \varepsilon_1$$

Adding all control variables (assuming a best fit with a quadratic term)

$$Y = \beta_0 + \beta_2 X_1 + \beta_3 X_1^2 + \varepsilon_1$$

$$Y = \beta_0 + \beta_2 X_1 + \beta_3 X_1^2 + \beta_2 X_{C1} + ... + \beta_{12} X_{C10} + \varepsilon_1$$

Testing for partial moderation, for each moderator

$$Y = \beta_0 + \beta_2 X_1 + \beta_2 X_1^2 + \beta_3 X_{C1} + ... + \beta_{13} X_{C10} + \beta_{14} X_m + \varepsilon_1$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_{C1} + ... + \beta_{13} X_{C10} + \beta_{14} X_m + +\beta_{15} X_1 X_m + \varepsilon_1$$

Y = Receipt of useful knowledge
X_1 = cohesion
$\beta_2 X_{C1} + ... + \beta_2 X_{Cn}$ = control variables
X_m = moderator being tested
ε_1 = error terms

Continuous variables were standardized to a normal distribution with mean one and standard deviation zero. This makes it possible to test mediation effects, and to compare the coefficient estimates (Hair et al., 2010).

Full-model testing
After hypotheses were tested, a full model for the receipt of useful knowledge was developed through step-wise estimation. Given the data, the goal was to specify the model with the highest predictive accuracy for the receipt of useful knowledge, while maintaining robustness and validity. The estimation was done following the Stepwise Estimation Method defined by (Hair et al., 2010), where the independent variable with the highest bi-variate correlation with the dependent variable is introduced first, and if the F-test is significant (rejecting the hypothesis that the next iteration is not a better model), the next variable is added, until no variable (combinations) can be found that result in a significant F-test.

3.6.3 Robustness testing

Several alternative models were developed in parallel of the OLS model, to evaluate the robustness of OLS findings and hypothesis tests. For robustness, five-fold cross-validated OLS, Weighted Least Squares (WLS) and (X-validated) and Mixed Linear Model (MLM) are used to test robustness. These three robustness methods are elaborated in this chapter.

Model stability testing
During the testing of hypotheses, and the specification of the full model, it seemed that the first hypothesis was quite strongly supported. In all hypothesis tests and the full model, cohesion and its quadratic were a significant contributor with p-values < 0.05 and $adj.R^2$ contributions. During robustness-checks, however, by randomizing the random train-test split, some coefficient estimates turned out to be very unreliable, including that of cohesion and its quadratic. At first

it was believed that the multivariate outlier identified in 4.4.1. was the culprit, but after further inspection, it turned out to not be the case, as shown in table 3.13.

Regression model tests[a]

df	R^2	$adj.R^2$	P-values			MD outlier
			Cohesion	Cohesion^2	Range alter	
22	0.402	0.310	0.04	0.08	0.17	present
22	0.370	0.274	0.12	0.20	0.23	present
22	0.466	0.385	0.08	0.14	0.87	present
22	0.454	0.371	0.10	0.16	0.91	present
22	0.416	0.326	0.12	0.06	0.40	present
22	0.362	0.265	0.53	0.88	0.18	not present
22	0.316	0.211	0.20	0.66	0.29	not present
22	0.328	0.226	0.03	0.39	0.38	not present
22	0.435	0.348	0.20	0.74	0.37	not present
22	0.353	0.254	0.20	0.78	0.30	not present
21	0.388	0.299	0.58	not included	0.58	no outlier
21	0.353	0.254	0.40	not included	0.37	no outlier
21	0.353	0.254	0.59	not included	0.60	no outlier
21	0.353	0.254	0.44	not included	0.30	no outlier
21	0.353	0.254	0.07	not included	0.12	no outlier

Table 3.13: Data instability

a) Fifteen runs of a semi-fully specified model. Random seed generator changed at every run. First five runs resulted in the MD 140 point present. Kept running until five models without the point were found to compare. Last five models without the quadratic of cohesion, but no check if the same point was present (if it was, it was no outlier anymore) to test whether instability was due to the outlier, the quadratic, or both.

The shown instability of coefficient estimate confidence might be due to influential datapoints (although they have been investigated in the chapters above) or due to the inherent amplified influence of individual datapoints due to the small dataset. As a result, 5-fold cross-validated results will be reported with the full model and hypothesis tests, to show which coefficient estimates are probably more robust to changing data. Additionally, a plot of the p-values of the five folds will be provided to show the sensitivity of p-value estimates to specific data-splits.

Weighted least-squares

As discussed in section 3.6.1, the chapter on regression assumptions, it was argued that observations were not independent, since respondents evaluated multiple interactions. Weighted least-squares is used as a robustness evaluation of the full model. Weighted regression adds weights to the observations in the regression model, in this case, putting 1/n weights on observations of which the respondent reported on n interactions.

This model does not solve the problem of independence of observation, because it introduces bias of its own. It operates under the assumption that the specified weights are correct, and rater-specific bias between observations scales with the amount of observations present by the rater. Furthermore, it assumes all interdependence effects equal, and amplifies any rater bias for respondents that reported less interactions.

Weighted regression is an evaluation of robustness however, because it shows the other end of two extremes, and as such, coefficient estimations that are significant for both the unweighted

OLS model and the weighted model can be more evaluated more confidently (although it does not guarantee validity).

Mixed linear model

The last robustness test included was the use of Mixed Linear Models (MLM), also referred to as Mixed models. MLM concern models with both fixed and random effects. Fixed effects are effects that are assumed to be without measurement error, with all possible occurrences present in the data (Kreft & De Leeuw, 1998). Gender or age are great examples of fixed effects. Random effects are effects that cannot be measured without error, given the nature of the study is a sample of the entire population (Kreft & De Leeuw, 1998). In this case, the respondent is a random effect, since it correlates observations amongst each-other from the same participant. MLM account for this random effect by estimating two types of variance; the variance between observations in the same group, and the variance that is not dependent on the group in which the observation is present (Simpson, 2001). Coefficient estimates are subsequently based on the fixed effects, not the random effects.

The random effect, in this instance, is the rater. The rater is the origin of all rater bias, interdependence between observations and controls for individual effects on the rater side of the dyad. These three points were identified in section 3.6.1 and section 3.5.3 to be the most problematic limitations to the interpretation of the coefficient estimates of the OLS model. MLM therefore offers a critical robustness evaluation, and significance in the MLM model signals stronger confidence in the contribution of a variable to the independent variable.

4. Results

In this chapter, the results of the study are presented. In section 4.1, the descriptives of the sample, the regression dataset, and the network data, are provided. Afterward, the correlation matrix of the regressiondata is provided in section 4.2. Subsequently, in the main part of this chapter, section 4.3 provides all regression model tests. The chapter closes with section 4.3.3, which discusses the robustness of the regression models provided.

4.1 Descriptive statistics

4.1.1 Sample description

For the survey study, 236 people were selected in the pre-defined network. In total, 170 people participated in the survey study. Out of these participants, 165 respondents were part of the predefined network, and could therefore be used for analysis. Four additional respondents are the result of the iterative improvement of the network definition over the course of the study, which led to the removal of these respondents from the pre-defined network. One excess participant resulted from an instance where a manager extended the survey invitation to a group with people not part of the pre-defined network.

These 165 participants evaluated 415 interactions. Because this study relies on complete interactions (both the ego and the alter participating in the study), many data-points were not suitable for regression analysis. Removing non-complete interactions resulted in 247 data-points. Of these interactions, 8 were removed because of self-reference, 7 were removed because of evidence of straight-lining, 21 were removed because of missing data, and 8 had to be removed because they contained outliers that had to be removed. The survey tool used could not measure completion time, but because of the significant length of the survey, it was assumed that controlling for straight-lining would remove impossibly fast responses. These filtering steps resulted in 203 observations, reported by 118 respondents, ready for regression analysis. A further description of the network data of the sample can be found in 5.1.3.

4.1.2 Regression descriptives

Of the 118 respondents, Average age was 36, and ranged between 22 and 60. 105 respondents were male (89%) and 13 female (11%). Of the respondents, 18 (15.3%) reported to have a managerial role (project lead or team manager). Average tenure was 7.4 years, ranging from 0 to 38, and with a median of 5, and 34 respondents indicating their tenure was 6 months or less, roughly following an exponential distribution. As elaborated in the introduction, the sample was quite heterogeneous in team-membership and location. Although it was clear from the start Dept. B was going to be a larger department, due to data removal of data, which has dis-proportionally hit groups with smaller numbers, has left some deparments underrepresented. Unfortunately, findings on differences between departments, can therefore not be generalized. In table 4.1, percentages are presented, compared to the sample and the defined network.

Network variables cohesion and network diversity roughly followed power-law distributions, where lower values were much more frequent than higher values. Tie strength seems more normally distributed, but highly skewed to lower scores. Of the ties evaluated, 23 (11.3%) were between

Department	Number of resp.	% of resp.
Department A	3	2.5%
Department B	87	73.7%
Department C	5	4.2%
Department D	15	12.7%
Other department	4	3.4%
Other department	3	2.5%
Location		
Europe	60	50.8%
The U.S.A.	36	30.5%
Asia	22	18.7%

Table 4.1: Distribution of respondents across departments and locations

friends. As the friendship-dummy was uncorrelated with tie-strength, this in-balance is not deemed to be problematic.

Likert-scale variables Tacitness and Receipt of useful knowledge both fit normal distributions, although the receipt of useful knowledge was skewed to the positive, with a mean score of 5.05 out of 7.

Most interactions were reported on colleagues that were close, both cognitively as well as geographically. For geographical proximity, 73% of interactions were reported in the same building, and for cognitive distance, 41.7% of the interactions were both in the same department and with the same functional background. The distributions are shown in table 4.2.

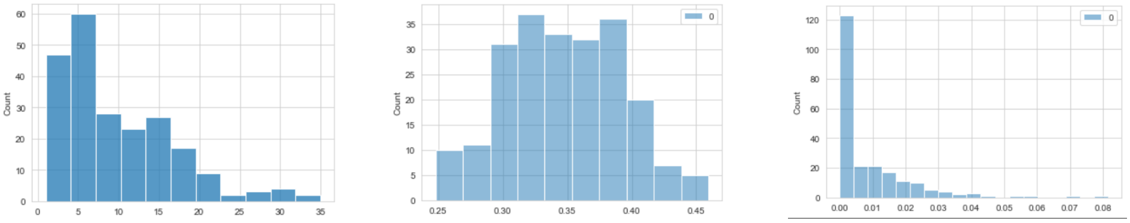
Geopraphical proximity	Number of int.	%of int.
The same floor	128	62.7%
The same buidling	21	10.3 %
Different building	12	5.9 %
Different site	43	21.1 %
Cognitive distance		
Same func. background and department	85	41.7 %
Same func. background or department	90	44.1 %
Neither	29	14.2 %

Table 4.2: Geographical and congitive distance in the interaction dataset

4.1.3 Network descriptives

For the bounded network, participants identified 1107 knowledge-seeking relationships. Only 222 people in the bounded network of 236 employees appeared as nodes in the network. The network data describes a clustered network, centralized around a few important brokers. Network visualizations additionally show that some groups are more central than others. As the histograms below in figure 4.1 show, average distance and network size are proportionally distributed, but betweenness-centrality is skewed to a smaller group of influential individuals. This suggests that enabling efficient knowledge flow, especially through those key brokers, is highly important.

Network metrics are studied in more detail to reveal the nature of the network. Table 4.3 shows descriptive information on the structure of the network, compared to a random network.



(a) Histogram of networksize of nodes in the network
(b) Histogram of closeness centrality of nodes
(c) Histogram of betweenness-centrality of nodes

Figure 4.1: Histograms of the bounded network

From this comparison, conclusions can be drawn on its composition.

Bounded network	**Value**
Number of nodes	222
Number of edges	1107
Average clustering	0.30
Average geodesic	2.93
Coefficient Sigma	2.545
Coefficient Omega	0.345
Random Graph[a]	
Random average clustering	0.04
Random average geodesic	2.63

Table 4.3: Network descriptives

a) Random graph based on n=222 and p = 0.0449, the average probability of connection in the bounded graph (average degree divided by the amount of nodes). Random graph shown is reference, coefficients were computed with ten randomly generated graphs.

According to Watts and Strogatz (1998) and Humphries and Gurney (2008), the network shows a small-world nature, where the clustering and average path length are both higher than the random-graph, but the difference between clustering and random clustering is much higher than between path length and random path length. To verify, coefficient sigma (Humphries & Gurney, 2008) and coefficient omega (Telesford et al., 2011) are computed. A sigma above 1, and an Omega around 0, indicate that indeed the network inhibits small-world dynamics. This is to be expected in a social network, and gives confidence in the validity of the network variables.

Network visualization, combined with contents of the exploratory interviews, were used to check whether network graphs represented the sample. In discussions with quality management, continuous improvement, knowledge management, and department stakeholders, remarks of recognition of network groups and interconnections, as well as surprise to which some groups were connected, further validated the network. The graph shown in figure 4.2 represents the different departments of the usecase, and showing in-betweenness centrality of nodes.

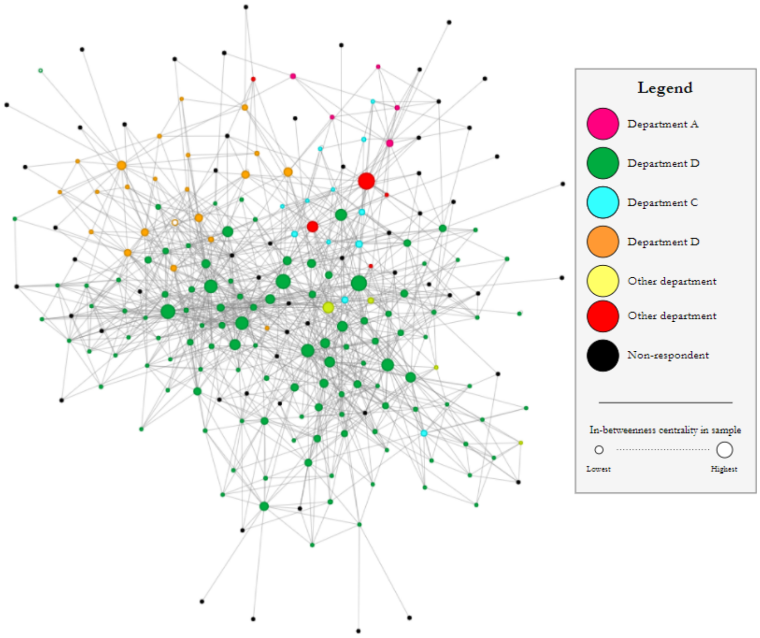


Figure 4.2: Graph of the bounded network
Picture was constructed using a Barnes-Hut algorithm, using Pyvis and NetworkX extentions of Python

4.2 Correlation matrix

The correlation matrix is used for two purposes; firstly, it is needed to screen for bivariate collinearity. High correlations between independent variables can make the estimation of coefficients and standard deviation in regression analysis problematic (Hair et al., 2010). Secondly, it is used to evaluate bivariate correlations related to the hypotheses, as well as screening for potential candidates for the full-model. Step-wise addition, used for the full model, occurs based on bi-variate correlations with the dependent variable (Hair et al., 2010).

The the correlation matrix depicted in table 4.4 shows no evidence for problematic bi-variate or multivariate correlation. This is in line with the multicollineariy checks performed in section 3.6.1. Strong correlations only occur between tenure and age of the alter (0.654, $p < 0.05$) and the ego (0.576, $p < 0.05$), but this is expected given their clear causal relationship.

For RUK, complexity (0.243, $p < 0.05$), tacitness (-0.363, $p < 0.05$), manager ego (-0.231, $p < 0.05$) and alter (-0.161, $p < 0.05$), tie-strength (-0.224, $p < 0.05$), cognitive distance (-0.160, $p < 0.05$) and proximity (-0.230, $p < 0.05$) are correlated. Since this study evaluates the interaction effects between cohesion and other antecedents of RUK, these are also evaluated. Cohesion shows bi-variate correlation with age alter (-0.132, $p < 0.05$), manager alter (-0.141, $p < 0.05$), proximity (-0.176, $p < 0.05$) and the range of the ego (-0.152, $p < 0.05$) and alter (-0.185, $p < 0.05$). Other higher correlations that stand out are between managerial role and range (0.359, $p < 0.05$ for the ego, 0.461, $p < 0.05$ for the alter) and between the range of ego and alter (0.387, $p < 0.05$), but these too seem explainable and fall within expected bounds of what can be expected.

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. Receipt of useful knowledge (RUK)	5.03	1.14									
2. Complexity	3.34	0.73	**0.243***								
3. Tacitness	4.41	1.50	**-0.363***	-0.145							
4. Age ego	37.06	8.64	-0.083	0.055*	-0.032						
5. Gender ego	0.89	0.31	-0.027	0.011	0	0.042					
6. Manager ego (dummy)	0.18	0.38	**-0.231***	-0.081	0.076	**0.233***	0.001				
7. Friends (dummy)	0.11	0.32	0.111	-0.004	-0.056	0.036	-0.068	-0.047			
8. Cohesion	0.02	0.02	0.129	-0.045	-0.057	-0.049	-0.012	-0.084	0.013		
9. Tenure ego	8.23	9.02	-0.01	0.088	0.095	**0.576***	0.118	**0.208***	-0.02	-0.102	
10. Age alter	39.81	8.93	-0.083	**0.181***	0.097	**0.14***	-0.007	0.072	-0.068	**-0.132***	0.123
11. Gender alter	0.89	0.31	-0.029	0.035	-0.109	**0.054***	0.075	0.12*	-0.116	-0.008	0.033
12. Manager alter (dummy)	0.19	0.40	**-0.161***	-0.1	0.102	0.057	0.055	**0.387***	**-0.136***	**-0.141***	0.016
13. Networksize ego	10.79	4.74	-0.119	0.113	-0.027	0.163	0.114	0.048	-0.037	-0.107	0.069
14. Networksize alter	10.13	5.19	0.072	0.011	-0.114	**0.05***	0.003	-0.115	-0.054	-0.006	0.124
15. Tenure alter	11.68	11.30	-0.132	0.113	0.058	0.166	-0.057	**0.201***	-0.052	-0.106	0.097
16. Tie-strength	0.14	0.09	**0.224***	0.017	-0.057	-0.097	-0.062	-0.078	0.121	0.098	-0.083
17. Cognitive distance	0.75	0.71	**-0.16***	-0.091	-0.066	**0.153***	0.173*	**0.288***	-0.06	-0.103	**0.197***
18. Educational dissimilarity	1.61	1.43	0.081	0.121	-0.015	0.04	-0.052	-0.044	-0.027	-0.048	0.022
19. Physical distance	0.90	1.25	**-0.23***	0.103	0.1	**0.257***	-0.004	**0.134***	0.005	**-0.176***	0.13
20. Expertise overlap	0.27	0.37	0.073	**0.154***	**-0.193***	-0.12	-0.01	-0.134	0.11	0.073	**-0.133***
21. Range ego	0.28	0.23	-0.078	-0.02	**-0.188***	0.108	0.208*	**0.359***	0.038	**-0.152***	0.074
22. Range alter	0.35	0.28	0.038	0.084	-0.109	**0.15***	0.043	**0.239***	0.009	**-0.185***	0.057

Variable	10	11	12	13	14	15	16	17	18	19	20	21
11. Gender alter	**0.292***											
12. Manager alter	0.017	**0.169***										
13. Networksize ego	0.096	0.095	-0.102									
14. Networksize alter	0.076	**0.194***	**-0.146***	-0.029								
15. Tenure alter	**0.654***	**0.245***	**0.143***	-0.107	0.025							
16. Tie-strength	-0.095	-0.096	-0.025	**-0.691***	-0.01	0.013						
17. Cognitive distance	-0.063	0.088	**0.194***	0.045	0.005	-0.014	-0.097					
18. Educational dissimilarity	0.035	-0.041	0.06	-0.073	0.037	-0.017	0.13	-0.126				
19. Ordinal proximity	0.015	0.055	**0.133***	**0.178***	0.055	0.011	**-0.206***	**0.28***	-0.022			
20. Expertise overlap	-0.01	**0.137***	0.053	0.001	-0.012	-0.011	-0.06	**-0.2***	0.115	0.104		
21. Range ego	-0.066	**0.221***	**0.237***	0.082	0.055	0.007	**-0.166***	**0.371***	-0.105	**0.298***	0.058	
22. Range alter	0.091	**0.241***	**0.461***	0.044	**-0.204***	0.031	0.033	**0.277***	**0.137***	0.102	0.075	**0.387***

*$p < 0.05$

Table 4.4: Correlation matrix

4.3 Regression analysis

Hypothesis test results, tested in stepwise-built models, are discussed below. All continuous variables were standardized before analysis to make coefficients more interpretive. Results are accompanied with R^2 and $adjustedR^2$ scores. Both measures indicate how much of the variance of the dependent variable is explained by the independent variables in the model, but $adjustedR^2$ includes the number of variables in the calculation (Hair et al., 2010). In this way, it penalizes variables that do not add anything to the explained variance, therefore decreasing the $adjustedR^2$ score when comparing step-wise built models where only one variable is added-removed.

4.3.1 Hypothesis tests

Table 4.5 shows that the quadratic term of bounded cohesion is not significant (β = -0.0497, p > 0.05), and neither is bounded cohesion itself (β = -0.208, p > 0.05). The quadratic term is also not significant in both the WLS model (β = 0.0036, p > 0.05) and the X-validated model (β = -0.055, p > 0.05). This rejects hypothesis 1. Table 4.6 shows that tie-strength is positively related to the receipt of useful knowledge (β = 0.2815, p < 0.05), but the moderator between cohesion and tie-strength is not, providing no support for hypothesis 2. Interestingly, the network diversity of the receiver is not significant in relation to RUK (β = -0.0459, p > 0.05), but the moderator of network diversity on cohesion, is positively related to RUK (β = 0.2099, p < 0.05). The WLS model also shows a similar positive significant effect (β = 0.2869, p < 0.05), but the cross-validated model does not (β = 0.2068, p > 0.05). In table 4.7, the network diversity of the alter, does not have a significant relationship with RUK (β = 0.101, p > 0.05). Additionally, the moderating effect between range alter and cohesion is also non-significant (β = -0.116, p > 0.05). As a result, hypothesis 4a has to be rejected, instead, effect seems to be significant in the opposite direction. There is no support for hypothesis 4b. For comparison, tacitness and its interaction effect are also included, since complexity could not be tested. There is a strong, significant, negative relationship between tacitness and RUK (β = -0.4158, p < 0.001), but its moderating effect with cohesion is also non-significant (β = -0.0299, p > 0.05).

In table Table 4.7, two remaining interaction effects are tested. First of all, proximity has a positive relationship with RUK (β = -0.012, p > 0.05), but its moderation effect with cohesion is insignificant (β = -0.082, p > 0.05), rejecting hypothesis 3. Congitive distance is on its own not related to RUK, but the multiplication between cohesion and cognitive distance is (β = 0.223, p < 0.1). As a result, there seems to be weak evidence of mediation. The moderation between cohesion and range is significant for the WLS model (β = -0.395, p < 0.05), but not for the five-fold cross-validated model (β = -0.239, p > 0.05) between Hence, Hypothesis 5 is not rejected.

4.3.2 Full-model

To investigate the predictive value of the independent variables, a full regression model is built, based on the Stepwise Estimation Method by (Hair et al., 2010) and theory. The model is built based on the biggest bi-variate correlations between the receipt of useful knowledge and the independent variables. At every step, the F-statistic was computed, to indicate whether the addition of the extra independent variable was a significant improvement over the last model.

As seen in the full model, step-wise addition leads to the inclusion of tacitness, proximity, tie-

DV: Receipt of useful knowledge (RUK)

Variable	Model 1	Model 2	WLS	X-validated	MLM
const	4.957	5.011	4.702	5.200	4.911
Gender ego	-0.045	-0.034	0.474*	-0.022	0.007
Gender alter	0.214	0.211	0.232	0.215	0.264
Manager ego	-0.623**	-0.631***	-0.165	-0.620*	-0.681***
Manger alter	-0.077	-0.077	0.121	-0.079	-0.108
Friends	0.376	0.341	0.440	0.348	0.421
Age ego	-0.096	-0.063	-0.020	-0.019	-0.093
Age alter	0.040	0.036	-0.236**	-0.063	0.026
Networksize ego	-0.124	-0.142	0.093	0.033	-0.108
Networksize alter	0.024	0.004	-0.015	-0.143	0.030
Tenure ego	0.097	0.073	0.264***	0.001	0.139
Tenure alter	-0.178	-0.159	0.137	0.072	-0.176
Cohesion	0.101	0.208	0.256**	0.211	0.104
Cohesion^2		-0.050	0.003	-0.055	0.009
R^2	0.116	0.128	0.188	0.145	—
adj-R^2	0.062	0.069	0.119	0.072	—

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.5: Cohesion hypothesis test

DV: Receipt of useful knowledge (RUK)

Variable	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	WLS	X-valid.	MLM
const	5.075	5.073	5.196	5.218	5.010	5.041	4.805	5.172	4.977
Gender ego	-0.068	-0.067	-0.050	-0.060	-0.033	-0.029	0.627**	0.007	0.015
Gender alter	0.174	0.174	0.000	-0.012	0.211	0.181	0.161	0.217	0.231
Manager ego	-0.614**	-0.614**	-0.590***	-0.589***	-0.631**	-0.577**	-0.495*	-0.545	-0.624**
Manger alter	-0.025	-0.025	0.009	0.010	-0.077	-0.085	0.134	-0.070	-0.114
Friends	0.301	0.301	0.315	0.302	0.342	0.298	0.138	0.319	0.358
Age ego	-0.090	-0.090	-0.128	-0.128	-0.063	-0.083	-0.022	-0.020	-0.102
Age alter	0.021	0.022	0.090	0.092	0.036	0.027	-0.162	-0.075	0.011
Networksize ego	0.070	0.069	-0.153*	-0.154*	-0.142*	-0.156*	-0.075	0.020	-0.131
Networksize alter	0.019	0.019	-0.034	-0.035	0.004	-0.007	-0.091	-0.155	0.011
Tenure ego	0.080	0.080	0.149	0.152	0.073	0.099	0.188*	-0.003	0.175
Tenure alter	-0.113	-0.114	-0.145	-0.146	-0.159	-0.156	0.263	0.097	-0.165
Cohesion	0.212*	0.213*	0.196*	0.200*	0.208*	0.231**	0.234*	0.224	0.178**
Cohesion^2	-0.063	-0.063	-0.057	-0.062	-0.050	-0.031	0.029	-0.029	-0.002
Tie-strength	0.282**	0.280**							
Tacitness			-0.411***	-0.416***					
Range ego					-0.001	0.046			0.044
Cohesion x Tie-strength		-0.004							
Cohesion x Tacitness				-0.030					
Cohesion x Range ego						0.210**	0.287***	0.207*	0.249***
R^2	0.155	0.155	0.244	0.245	0.128	0.150	0.236	0.156	—
adj-R^2	0.094	0.089	0.189	0.186	0.064	0.084	0.166	0.084	—

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.6: Hypothesis tests tie-strength, Tacitness and Range ego

DV: Receipt of useful knowledge (RUK)

Variable	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	WLS	X-valid.	MLM
const	5.108	5.106	5.137	5.129	5.037	4.946	4.667	5.130	4.898
Gender ego	-0.025	-0.026	-0.066	-0.072	0.014	0.080	0.589**	0.049	0.104
Gender alter	0.191	0.171	0.239	0.233	0.228	0.240	0.293	0.221	0.276
Manager ego	-0.673***	-0.659***	-0.626***	-0.619***	-0.570**	-0.543**	-0.758**	-0.597*	-0.591**
Manger alter	-0.129	-0.131	0.013	0.043	-0.063	-0.053	0.039	-0.057	-0.082
Friends	0.338	0.340	0.359	0.377	0.331	0.379	0.536	0.405	0.446*
Age ego	-0.072	-0.067	-0.184	-0.029	-0.058	-0.083	-0.016	-0.021	-0.108
Age alter	0.028	0.019	0.038	0.030	0.026	0.014	-0.175	-0.086	0.005
Networksize ego	-0.130	-0.132	-0.109	-0.114	-0.142*	-0.142*	-0.021	0.022	-0.113
Networksize alter	0.024	0.042	0.023	0.017	0.007	-0.004	-0.059	-0.142	0.019
Tenure ego	0.078	0.062	0.062	0.061	0.079	0.068	0.176*	-0.005	0.135
Tenure alter	-0.147	-0.143	-0.194*	-0.202*	-0.160	-0.171	0.229*	0.062	-0.181
Cohesion	0.222*	0.134	0.149	0.089	0.195*	0.015	-0.049	0.020	-0.023
Cohesion	-0.048	-0.055	-0.030	-0.020	-0.047	-0.022	0.025	-0.023	0.013
Range alter	0.101	0.092							
Proximity			0.012***	0.163***					
Cognitive distance					-0.133	-0.124			-0.136
Cohesion x Range alter		-0.116							
Cohesion x Proximity				-0.082					
Cohesion x Cog. distance						0.233*	0.395***	0.239	0.161
R^2	0.132	0.137	0.162	0.166	0.133	0.148	0.245	0.156	—
adj-R^2	0.069	0.070	0.101	0.100	0.070	0.081	0.176	0.077	—

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.7: Hypothesis tests range alter, geographical proximity and cognitive distance

DV: Receipt of useful knowledge (RUK)

Variable	Model 15	Model 16	Model 17	Model 18	Model 19	WLS	X-valid.	MLM
const	4.941***	5.120***	5.310***	5.317***	5.292***	5.042***	5.286***	5.279***
Gender ego	-0.023	-0.033	-0.077	-0.091	-0.057	0.450	-0.050	-0.033
Gender alter	0.266	0.051	0.057	0.042	0.015	-0.169	0.019	0.012
Manager ego	-0.657***	-0.617***	-0.604***	-0.593***	-0.530**	-0.497**	-0.525**	-0.530**
Manager alter	-0.135	-0.048	0.079	0.098	0.106	0.179	0.105	0.100
Friends dummy	0.337	0.307	0.326	0.291	0.267	0.101	0.271	0.272
Age ego	-0.087	-0.148	-0.088	-0.106	-0.126	-0.047	-0.128	-0.121
Age alter	0.034	0.091	0.088	0.072	0.067	0.116	0.063	0.060
Networksize ego	-0.120	-0.129*	-0.097	0.067	0.040	0.147	0.039	0.059
Networksize alter	0.031	-0.008	0.008	0.022	0.010	0.016	0.016	0.007
Tenure ego	0.095	0.172*	0.163*	0.161*	0.186**	0.182**	0.187*	0.184**
Tenure alter	-0.193	-0.177	-0.201	-0.161	-0.144	-0.094	-0.146	-0.138
Tacitness		-0.430***	-0.447***	-0.432***	-0.433***	-0.399***	-0.431***	-0.430***
Physical distance			-0.219***	-0.199***	-0.162**	-0.173**	-0.164**	-0.163**
Tie strength				0.224**	0.199*	0.205**	0.200	0.218**
Cohesion					0.125	0.159**	0.123	0.124
Cohesion x range ego					0.165*	0.132*	0.163*	0.158*
R^2	0.114	0.241	0.289	0.307	0.326	0.338	0.339	—
adj-R^2	0.063	0.194	0.241	0.256	0.268	0.281	0.266	—
F-statistic	—	32.044***	12.872***	4.800**	2.563*	—	—	—

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.8: Full model

strength and the pair-wise addition of log-transformed cohesion and the moderator between ego network diversity and cohesion. To make sure this model was not a local optimum, a step back was made, and the model was tested separate of cohesion and the moderator and from tie-strength. With these models, all independent variables were added one by one based on their correlation with the receipt of useful knowledge, but non of the additions were able to achieve a statistically significant F-statistic.

4.3.3 Robustness

To test the robustness of findings, two comparative measures were used. Five-fold cross-validated models were used to produce less data-sensitive estimates, and weighted-least squares models were used to produce estimates that were less dependent on respondents with multiple interactions. The rationale of these methods is discussed in section 4.6.2.

Five-fold cross validation

Cross-validated estimates show that the estimations for Tacitness and Proximity are the most robust across different splits of data. Tie strength, estimates for Cohesion and the moderator of Cohesion and Range ego show to be less robust, and control-variables in general move away from significance. In figure 4.3, p-value estimates for model variables over five folds of data is shown, to show the sensitivity of variables to specific data splits. The plot shows significant model instability, suggesting that the estimations of beta coefficients are significantly impacted by the limited sample size. Implications are discussed in section 5.2.4.

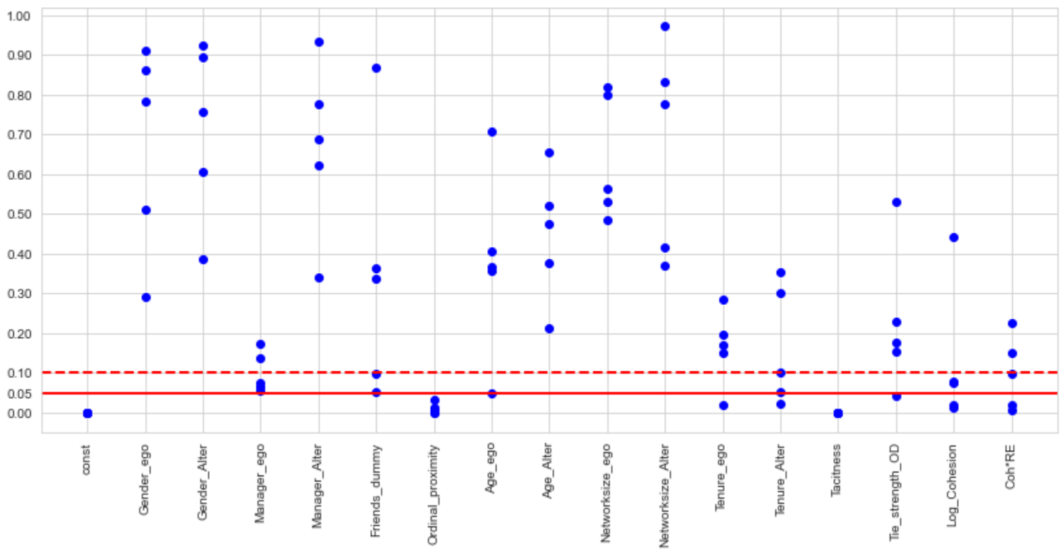


Figure 4.3: Five-fold cross validated p-value estimates for the variables in the final model
The data was split in five parts, after which four parts were used for training, and one used for testing. The resulting p-value estimates are provided, depicted with five dots per variable. The cross-validated estimates in general result in somewhat higher p-values, as there is less training data in each model, and the spread of the dots for each variable shows the general instability of the model in general, and the sensitivity of the data for the estimation of specific variables to specific data-splits.

Weighted regression

Weighted regression results are shown next to the final model, in the results chapter of this study. These results show coefficient estimates that put more weight on observations of participants with only one interaction, reducing the impact of rater bias of multiple-raters, and increasing the bias of single-raters. Coefficient estimates significant for both models show more robustness to rater-specific bias.

Results are good, showing that Tacitness, Proximity and Tie-strength, variables that are more rater dependent, are robust in WLS models. Control variables are, in general, less robust, which was to be expected. Range ego and Cohesion have diminished exposure to rater bias in general, since the metrics are made up of observations by multiple respondents, and hence Cohesion and the moderator therefore do not have to specifically significant in this model in order to be robust.

How the results of this study relate to its hypotheses is summarized in table 4.9 below.

	Hypothesis	**Support**
H1	Cohesion has an inverted U-shape relationship with the receipt of useful knowledge, such that the quadratic has a negative significant coefficient.	Not supported
H2	Tie Strength moderates the relationship between cohesion and the receipt of useful knowledge, such that an increase in tie strength lowers the optimal level of network cohesion for receipt of useful knowledge.	Not supported
H3	Physical distance moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in distance increases the optimal degree of network cohesion for receipt of useful knowledge	Not supported
H4a	The range of the receiver moderates the relationship between cohesion and receipt of knowledge transfer, such that it decreases the optimal degree of network cohesion for the receipt of useful knowledge.	Not supported, opposite effect partially supported
H4b	The range of the sender moderates the relationship between cohesion and receipt of knowledge transfer, such that it decreases the optimal degree of network cohesion for the receipt of useful knowledge.	Not supported
H5	Cognitive distance moderates the relationship between cohesion and receipt of useful knowledge, such that an increase in cognitive distance increases the optimal degree of network cohesion for receipt of useful knowledge.	Weak support
H6	Complexity moderates the relationship between cohesion and receipt of knowledge transfer, such that an increase in complexity increases the optimal degree of network cohesion for receipt of useful knowledge	Could not be tested

Table 4.9: Summary of hypothesis test results

4.4 Post-Hoc analyses

Next to the hypothesis tests and the full model specification, several post-hoc analyses were conducted. This chapter presents post-hoc analyses on using MLM to test dyadic variables, dissecting physical distance and the most significant contributors to its relationship with RUK, and an investigation in what factors help alleviate the impact of Tacitness on RUK.

4.4.1 Post-hoc 1: Dyadic control variables

In the main models presented in section 4.3.1, section 4.3.2 and section 4.3.3, most control variables were operationalized on the individual level, and provided for both the ego and the alter. This turned out to be insight full, as, even in WLS and MLM, some of these variables showed significant difference between ego and alter, contributing to the dyadic knowledge flow narrative. Most studies on dyadic knowledge flow, such as (Reagans & McEvily, 2003; Levin & Cross, 2004; Tortoriello et al., 2012), control for individual differences, but do not test for them, only using dyadic variables in their analyses. Mirroring findings in this study to these seminal papers is restrained, however, as different control variables can lead to different results. To aid generalization, this post-hoc replicates the full-model with MLM, using dyadic variables, and controlling for individual effects. table 4.10 shows the results of this investigation.

DV: Receipt of useful knowledge (RUK)[a]

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
const	4.665***	4.863***	5.035***	4.982***	4.987***	5.014***	5.133***
Different gender	0.094	-0.046	0.013	0.049	0.057	0.039	0.041
Different role	0.337	0.234	0.230	0.229	0.201	0.193	0.061
Friends	0.429	0.404	0.387	0.328	0.324	0.282	0.306
Age difference	-0.060	-0.055	-0.066	-0.021	-0.022	-0.036	-0.016
Networksize difference	0.078	0.057	0.050	0.070	0.088	0.089	0.100
Tenure difference	-0.017	0.017	-0.036	-0.047	-0.04	-0.033	-0.048
Tacitness		-0.421***	-0.439***	-0.423***	-0.418***	-0.413***	-0.399***
Physical distance			-0.253***	-0.218***	-0.205***	-0.175***	-0.192***
Tie strength				0.215***	0.214***	0.203***	0.219***
Cohesion					0.081	0.140*	0.508**
Cohesion x range ego						0.189**	
Cohesion x DR							-0.468*

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

a) Mixed linear model, with ego (prespondent) as random effect. Min group size 1, max group size 4, group size average 1.7

Table 4.10: Full-model in MLM, with dyadic control variables

The results indicate robustness for dyadic variables. In general, dyadic variables seem to add less to explained variance, given higher significance levels for the significant variables in the full model. This cannot be concluded with certainty, however, as R^2 measures are missing. In model 4, tacitness (β = -0.423, p < 0.01), proximity (β = -0.218, p < 0.01) and tie strength (β = -0.215, p < 0.01) highly significant, and they are also consistently highly significant in other models. Findings for the interaction effect (β = 0.189, p < 0.05) between cohesion and range ego are also replicated in model 6. Additionally, model 7 shows that one of the dyadic control variables, different role, is weakly significant through its interaction with cohesion (β = -0.468, p < 0.10), increasing the significance of cohesion itself (β = 0.508, p < 0.05). This effect was uncovered while investigating tacitness, and is discussed further in section 4.4.3, and only added here to make its role comparable to other effects tested.

4.4.2 Post-hoc 2: Deconstructing physical distance

The full-model presented in section 4.3.2 showed that physical distance has a strong, negative relationship with RUK, where colleagues at further physical distance report lower scores for the receipt of useful knowledge. Physical distance in this model is ordinal, where a score between 0 and 3 represents to what extend ego and alter are present on the same (= 0) or a different floor (= 1), a different building (= 2) in the same municipality, or on a different continent (= 3)[1]. This ordinal scale was developed based on context, and inspired by (Levin & Cross, 2004), but does not specify where exactly these distances matter most. As a result, an investigation is conducted in several aspects of distance; the effect of being on a different floor in the same building, in a different building in the same municipality, between which continents the distance is measured, and directionality; whether distance works the same way in both directions.

DV: Receipt of useful knowledge (RUK)

Variable	**Model 1** *(N = 204)*	**Model 2** *(N = 147)*	**Model 3** *(N = 161)*
const	5.104***	4.953***	4.871***
Gender ego	-0.032	0.095	0.123
Gender alter	0.029	0.241	0.288
Manager ego	-0.091***	-0.505*	-0.649***
Manager alter	0.047	0.075	0.087
Friends	0.084	0.389	0.349
Age ego	-0.214	-0.116	-0.085
Age alter	-0.061	0.224	0.098
Networksize ego	0.279	-0.124	-0.098
Networksize alter	-0.647	0.042	0.034
Tenure ego	-0.025	0.101	0.053
Tenure alter	0.355**	-0.412***	-0.291***
Physical distance	-0.195***		
Different floor		-0.289	
Different building			-0.354
R^2	0.152	0.166	0.117
adj-R^2	0.099	0.092	0.046

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.11: Different types of physical distance and their impact on RUK

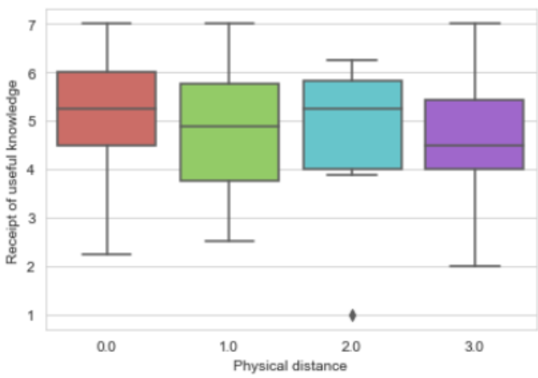


Figure 4.4: Physical distance and the receipt of useful knowledge values for samples of knowledge flow across different geographical borders

[1]in the sample, there were no instances where people worked with others in the same country, but on a different campus, and no instances where people worked with others in a different country, but on the same continent

Sample data: Receipt of useful knowledge (RUK)				
T-test	**Sample 1**	**Sample 2**	**Statistic**	***sig.***
1	EU ↔ USA	USA ↔ Asia	-1.216	0.233
2	EU ↔ USA	EU ↔ Asia	-0.074	0.942
3	EU ↔ Asia	USA ↔ Asia	-0.851	0.404
4	EU → Asia	Asia ← EU	-2.024	0.082*
5	EU → USA	USA ← EU	0.794	0.438
6	USA → Asia	Asia ← USA	-0.097	0.924

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.12: Student t-tests for the receipt of useful knowledge

Firstly, the presence of colleagues on the same, or a different floor, is investigated. Data was split to only represent interactions taking place in the same building. Model 2 in table 4.11 shows no significant effect for colleagues working on the same floor in the same building (β = -0.289, p > 0.10). In model 3, a new dummy variable, whether people working in the same municipality also work in the same building (0 for same building, 1 for different building), also shows no significant effect (β = -0.354, p > 0.10). Figure 4.4, a box-plot of the different values for physical distance, shows where the relative low predictive value of these variables comes from; most difference in the receipt of useful knowledge occurs between continents, with significant time-differences. As a result, differences between inter-continental knowledge flow was investigated further.

Tests for continental effects were split up into to parts. One part tested for significant differences in exchanges between colleagues working from specific continents. Here, both knowledge flow from and to the U.S.A. and the EU, for instance, were counted towards the same group. The second part tested directionality; here, differences between knowledge flows from and to continents were tested, so see whether RUK-scores show significant differences between Asia to EU and EU to Asia, for instance. To tests significant differences between these samples of interactions, student t-tests were conducted, to test the hypothesis H_0, that both groups represent a sample drawn from a distribution with the same mean. Table 4.12 shows the results of these tests; all but one test were insignificant, with only the directional difference between the EU and Asia reporting a weakly significant effect (T-test statistic = -2.024, p = 0.082). Insignificant effects could be due to two factors; both directionality and continental differences, if they exist, are far less significant than the difference between intra- and intercontinental knowledge flow. Alternatively, results were insignificant because of the very small sample of intercontinental knowledge flow; the number of intercontinental knowledge flows totalled 49, and all splits between groups within inter-continental knowledge flows are smaller still, bi-directional flows between continents ranging from 9 to 19 interactions, and unidirectional samples ranging from 4 to 10 interactions. Subsequently, results remain inconclusive.

4.4.3 Post-hoc 3: Breaking down Tacitness

The full-model presented in section 4.3.2 showed that Tacitness was the strongest predictor of the perceived usefulness of knowledge by the receiver. This finding has a number of important implications for practice, and further investigation into what makes perceived tacitness occur in knowledge-seeking interactions presents both scholars and practitioners with an opportunity for deeper understanding of its implications. Therefore, a post-hoc was conducted to study what variables predict Tacitness in regression. Two full-models were constructed, one in OLS, with individual control variables, and one in MLM, with dyadic control variables. Both models present significant variables relevant to Tacitness literature, and practitioners active within the context of this study.

Because this study has not fully established the appropriate control variables for a full model on Tacitness, and to fully utilize the data available, the same control variables were introduced for both models on Tacitness. This bolsters value of significant effects found, as these effects are significant despite a range of control variables, that might or might not be relevant for tacitness, correlating with the same DV. Furthermore, as the models show, some control variables play a role in the prediction of tacitness.

According to (Hair et al., 2010), full models are specified by adding the highest bi-variate correlated variables. In the OLS model, of all non-control variables, network diversity of the receiver was the only significantly correlated variable, and unsurprisingly significantly predicts tacitness in model 2 (β = -0.323, $p < 0.01$). Robustness checks show consistent significance of this variable across models. In step-wise addition and removal, two other weakly significant variables were found; the network diversity of the sender, and a moderation effect of tie strength on network diversity of the receiver. Range alter (β = -0.238, $p < 0.10$), and the pairwise addition of tie strength (β = -0.245, $p < 0.10$) and its interaction effect (β = -0.182, $p < 0.10$) in models 3 and 4 respectively, both fail to result in a significant F-statistic, almost reaching the $p < 0.05$ cut-off mark. The other structurally significant effects includes tenure of the receiver, significant in the first three models (β = 0.296, $p < 0.05$). Overall, the OLS models achieve low R^2 and *adj*-R^2 values, showing the limited predictive value of the variables included in the models.

For Tacitness, a dyadic level model, with MLM, was also specified. While this model does not include R^2 and *adj*-R^2 scores, severely limiting the evaluation of predictive power of the models, and hampering the comparison of models, the significant effects found still indicate which variables are most strongly related to tacitness. In the dyadic MLM model, previous findings were replicated first. In model 6, range ego was added, showing significance in this model as well (β = -0.189, $p < 0.05$). Range alter (β = -0.057, $p < 0.10$), and the pairwise addition of tie strength (β = -0.116, $p > 0.10$) and its interaction effect (β = -0.126, $p < 0.10$), parallel the findings in the OLS model. One thing that stands out in model 5 through 8, is the consistent significance of the different role variable, while this variable is not bivariately correlated with tacitness. After testing several interaction effects, it turned out that cohesion, while not significant on its own, is significant (β = -0.506, $p < 0.05$) when moderated by different role (β = 0.501, $p < 0.05$), in model 9. Model 9 shows that this is a moderating effect, as the relation between different role and tacitness becomes insignificant when the interaction with cohesion is introduced (β = -0.223, $p > 0.10$). This moderation effect is shown in figure 4.5. In the OLS and MLM models, the addition cohesion and tie strength on their own did not lead to significant effects. Other interaction effects were also tested, based on bi-variate correlations. Other control variables potentially relevant, such as the splitting of cognitive distance in functional background and departmental difference, did not result in significant effects.

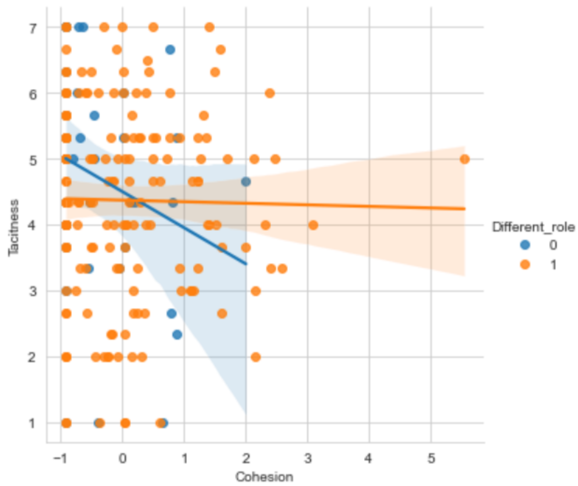


Figure 4.5: Different role moderates the relationship between cohesion and tacitness

DV: Tacitness

Variable	Model 1	Model 2	Model 3	Xval	WLS	MLM	Model 4	Xval	WLS	MLM
const	5.048***	4.617***	4.518***	4.517***	4.121***	4.456***	4.464***	4.464***	3.952***	4.411***
Gender ego	-0.037	0.170	0.116	0.128	0.017	0.100	0.236	0.214	0.139	0.157
Gender alter	-0.749**	-0.556	-0.439	-0.450	0.079	-0.370	-0.496	-0.472	0.075	-0.302
Manager ego	0.138	0.393	0.354	0.346	0.093	0.282	0.446	0.443	0.221	0.258
Manager alter	0.303	0.397	0.626*	0.620	0.534	0.684	0.390	0.392	0.407	0.274
Friends	-0.103	0.011	0.039	0.044	0.246	0.102	0.077	0.053	0.347	0.102
Age ego	-0.211*	-0.202	-0.176	-0.175	-0.147	-0.166	-0.187	-0.185	-0.144	-0.117
Age alter	0.200	0.150	0.200	0.202	0.021	0.212	0.137	0.128	0.002	0.094
Networksize ego	-0.031	-0.018	-0.021	-0.022	-0.075	-0.025	-0.176	-0.173	-0.216	-0.136
Networksize alter	-0.135	-0.105	-0.159	-0.161	-0.147	-0.168	-0.104	-0.107	-0.119	-0.074
Tenure ego	0.269**	0.252**	0.254**	0.256*	0.233*	0.241*	0.217*	0.220	0.182	0.137
Tenure alter	0.054	0.054	0.012	0.015	0.090	0.022	0.013	0.008	0.061	0.025
Range ego		-0.323***	-0.246**	-0.244	-0.319***	-0.234*	-0.373***	-0.371**	-0.415***	-0.247***
Range alter			-0.238*	-0.241	-0.143	-0.261**				
Tie strength							-0.245*	-0.246	-0.193	-0.178
Range ego x Tie strength							-0.182*	-0.188	-0.227	-0.104
R^2	0.087	0.122	0.138	—	—	—	0.146	—	—	—
adj-R^2	0.035	0.067	0.078	—	—	—	0.083	—	—	—
F-statistic	—	7.621***	3.342*	—	—	—	2.645*	—	—	—

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.13: Tacitness

DV: Tacitness[a]

Variable	Model 5	Model 6	Model 7	Model 8	Model 9
const	5.236***	5.185***	5.198***	5.152***	4.925***
Different gender	-0.502*	-0.231	-0.216	-0.225	-0.230
Different role	-0.502*	-0.397**	-0.420**	-0.402**	-0.223
Friends dummy	0.007	0.020	0.011	0.065	0.051
Age difference	0.043	0.013	0.020	-0.012	0.009
Size difference	-0.052	-0.027	-0.025	-0.041	-0.045
Tenure difference	0.095	0.046	0.041	0.056	0.051
Range ego		-0.189**	-0.169**	-0.214***	-0.184**
Range alter			-0.057*		
Tie strength				-0.116	
Tie strength x Range ego				-0.126*	
Cohesion					-0.506**
Cohesion x Different role					0.501**

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

a) Mixed linear model, with ego (prespondent) as random effect. Min group size 1, max group size 4, group size average 1.7

Table 4.14: Tacitness in dyadic control variables

5. Conclusion and discussion

This final chapter presents the conclusions of this study. First, in section 5.1 provides an overview of the most important conclusions, and provides an answer to the main research question. Afterward, the implications for theory are discussed in section 5.2.1, followed by implications for practice section 5.2.2, both addressing the first research sub-question. Subsequently, in section 5.2.3, recommendations for the use-case are provided, answering the second research sub-question. Following these sections, section 5.2.4 discusses the limitations of this study, and finally, section 5.2.5 discusses opportunities for future research.

5.1 Conclusion

This research set out to uncover how dyadic knowledge flows can be improved for the usecase of ASML. Embedded within a silo-ed context, and pressured by rapid growth and a decreasing relative number of senior engineers, this collection of heterogenous NPI teams interdependent on product knowledge faces the challenge of fostering effective and efficient knowledge flow across teams, departmentss and locations. To support the usecase, a survey study on its informal knowledge-seeking network was conducted. Its aim was to provide insight in how the receipt of useful knowledge could be improved by leveraging current strengths and elevating current weaknesses in the network. Towards this end, the following research question was formulated:

> *"To what extent do relational characteristics and knowledge characteristics impact the relation between structural network characteristics and the receipt of useful knowledge from dyadic knowledge flow, for the collaborative network of the usecase?"*

This study adopted the perspective of network friction and redundance. In order to make knowledge flow across dyads in organizational networks, relational, knowledge and structural characteristics of these networks interact to produce two opposing dynamics; alleviating friction, and introducing redundance. Relational, structural and knowledge characteristics can function as antecedents of successful knowledge flow (Szulanski, 1996; Ghosh & Rosenkopf, 2015), fostering trust (Levin & Cross, 2004), increasing transfer speed (Zander & Kogut, 1995) and enabling comprehension (Hansen, 1999; Sorenson et al., 2006).

These characteristics can also introduce redundancy, however, creating situations of decreasing access to new information and alternative perspectives. The time investment needed to maintain a excessively large and cohesive network limits informational access (Burt, 2004; Alguezaui & Filieri, 2010) and network redundance is related to a decrease in innovative capability (J.-C. Wang, Chiang & Lin, 2010; Tortoriello et al., 2015).

Ideally, contextual factors reduce friction enough to collaborative effectively, while not overcompensating and burdening overall exchange of knowledge, leading to inefficiencies. This study adopted the focus on cohesion, or triadic closure between dyads in collaboration networks, as the centre of investigation of this balance between friction and redundance. Accordingly, it was hypothesized that there would be a curvelinear relationship between cohesion and the receipt of useful knowledge.

Once the focus on triadic closure was established, five interaction hypotheses were drawn up. These centered around the idea that the other important contextual factors, cognitive distance, geographical proximity, network diversity, tie strength and knowledge complexity, would influence the relationship between cohesion and the receipt of useful knowledge, decreasing or increasing the need for cohesion to establish effective conduits for knowledge flow. With low complexity and low cognitive distance, knowledge should flow more easily, reducing the need for local triadic closure, and increasing the risk that two people have less to offer one-another in a clique. Adversely, with low tie strength, high geographical distance, and low network diversity, integrating and using knowledge becomes problematic, increasing the value of cohesion for the successful use of knowledge. Out of the five hypothesized relations, two were partially supported, either in both OLS and WLS (the moderation of network diversity on cohesion) or only in WLS (the moderation of cognitive distance on cohesion), resulting in partial support for the former, and weak support for the latter. The relationship with knowledge complexity could not be tested, unfortunately, given the invalidity of the measure, and Tacitness was tested instead, but returned no promising results.

These results indicated that the balance between friction and redundance cannot be found in a curvelineair effect between cohesion and effective knowledge flow, and that many of the contextual factors do not work interchangeably with cohesion to alleviate friction and introduce redundance, but instead a more nuanced perspective, where each contextual factor introduces a more separated effects, that can work in parallel to cohesion in the ability to enable or block dyadic knowledge flow. This study suggests that where cohesion is positive for the ease of knowledge transfer, when it comes to knowledge application, redundance introduced by cohesion makes the variable itself no longer contribute to the receipt of useful knowledge in cases of network diversity.

Instead, low tacitness, physical proximity, tie strength, the interaction between cohesion and network diversity, and the receiver not having a managerial role showed to be predictors of successful knowledge application. Tacitness was reduced in cases where the receiver had high network diversity, or where the two individuals in the dyad were both either manager or not a manager. Where they had a different role, it increased tacitness, unless sufficient cohesion was present between the two. For physical distance, knowledge application was only less successful when it concerned inter-continental knowledge exchange.

5.2 Discussion

5.2.1 Theoretical implications

In the pursuit of uncovering the mechanisms that drive dyadic knowledge flow, this study makes [number] of contributions to literature.

Firstly, as identified in the academic relevance chapter, the usecase presents a unique one in dyadic knowledge flow literature. While there are some imbalance problems in group distributions, findings that tacitness, proximity, tie-strength, cohesion and range playing a role in the receipt of useful knowledge, expand our understanding of under what conditions, and to what extent, these variables are significant predictors of successful knowledge application within an innovative network of product knowledge interdependent actors. This brief but consequential point attributes value to findings discussed in this chapter.

Secondly, this study nuances the discussion on network density and knowledge flow, as to the best of the authors knowledge, this is the first study to attempt to find a curvelineair relationship between cohesion and successful dyadic knowledge flow in a survey setting. As previously outlined, a curvelineair effect between cohesion the receipt of useful knowledge was a promising avenue of research, given the balance of friction and redundance in network structures (Ghosh & Rosenkopf, 2015; Burt, 2001), the prevalence of curvelineair effects in simulation studies (Cowan & Jonard, 2004; Zhao & Anand, 2013; Kim & Anand, 2018) and the gap of curvelineair effects in knowledge transfer literature (Wijk et al., 2008). This study failed to find any curvelineair effects for cohesion and receipt of useful knowledge, however. This finding has implications for literature. In studies of the ease of knowledge transfer, cohesion was shown to linearly alleviate friction (Reagans & McEvily, 2003). In simulation studies (Kim & Anand, 2018) and in studies in teams, curvilinear effects were found or proposed, that were thought to translate to dyadic effects. This study, instead, finds no curvilinear effect, but also no other direct relation between cohesion and the receipt of useful knowledge in dyadic knowledge exchange. Given the perspective of friction and redundance, this seems to suggests that where cohesion improves knowledge transfer-ability, this effect is nullified by redundance in its ability to create perceived value in knowledge application. Only in specific cases of friction between two actors, for instance when the receiver has a diverse network, or when the interaction takes place between people with different organizational roles, does cohesion contribute to knowledge application. In this role of alleviating friction, the interaction between cohesion and network diversity is consistent with literature, as (Reagans & McEvily, 2003) found network cohesion and network diversity to be complimentary in easing knowledge transfer. This suggests that previous survey studies were correct in the linear specification of cohesion, and the argument that cohesion and other contextual factors are complementary. Reagans and McEvily (2003) conclude that network diversity and network density are complementary, rather than opposing, for the ease of knowledge transfer. (Tortoriello et al., 2012) showed that both network diversity and cohesion are linearly and positively related to knowledge acquisition across organizational barriers. These findings originate from more homogeneous samples withing single departments, however, and this study argues that this is also the case for knowledge application in such a heterogeneous context of departments and locations, only the interaction between the two improve the receipt of useful knowledge, only either of the two does not contribute to knowledge application. This resonates with the view of (Burt, 2001) view of that a lack of both internal constraint (e.g. internal cohesion) and external constraint (e.g. external network diversity) represents a situation of optimal performance for a group. Tortoriello et al. (2015), in another study on knowledge application in patents how a balance between friction and redundance, where network redundance is negatively related, but the multiplication of redundance and diversity is positively related to innovative performance, suggesting a balancing effect through moderation instead. These findings suggests that, while cohesion improves knowledge dissemination, as knowledge transfer moves towards the application of knowledge, the effects of redundance become more pronounced, but not curvilinear, and cohesion on its own can have no effect, like in this study, or negative effects, like in (Tortoriello et al., 2015).

Thirdly, tacitness was found to be a very significantly negative predictor of the perceived usefulness of received knowledge. The important role of tacit knowledge it in its creation (Alavi & Leidner, 2001), diffusion (Hansen, 1999), sharing (Ganguly, Talukdar & Chatterjee, 2019) and use (Nakauchi et al., 2017) have been well established in literature. Despite suspicions that some of the strong correlation between tacitness and the receipt of useful knowledge is due to common method bias, due the degree of significance, confidence remains that tacitness plays a role in this usecase as well. Since sharing tacit knowledge is related to the innovative capability or an organization(Ganguly et al., 2019), understanding what makes tacit knowledge flow is valuable. In an additional post-hoc, this study attempted to estimate what variables related to tacitness, something less often quantitatively studied. It was found that network diversity of the receiver, role difference and the interaction between cohesion and role difference significantly and robustly related to tacitness.

Furthermore, the heterogeneity of the sample used for this study, in terms of geographical dispersion, presented an especially interesting finding in this study. Hypothesis tests, as well as the full-model, show that physical distance, an effect theorized to be of importance (Boschma, 2005; Kleinbaum et al., 2013), and shown to effect knowledge transfer between organizational units (Hansen & Løvås, 2004), but not controlled for in cohesion studies identified (Reagans & McEvily, 2003; Tortoriello & Krackhardt, 2010; Tortoriello et al., 2015; Levin et al., 2016). As (Kleinbaum et al., 2013) argue, geographic distance limits the ability to form ties, and tie-frequency is higher in employees working in the same office. These higher frequencies of interaction could explain a part of the impact cohesion has on succesful knowledge transfer, when not controlling, or using a sample without, geographical dispersion. this has two repercussions; while studies identified control for endogeneous effects, such as proximity, suitably, not controlling for geographic dispersion might still dis-balance our understanding of how knowledge flows across organizational boundaries. As (Phelps et al., 2012) argued, the role of geographical proximity in intra-organizational and interpersonal knowledge adoption is currently not well understood. As the receipt of useful knowledge has been identified to be a measure of knowledge application, this study contributes to the gap identified by (Phelps et al., 2012), that extreme distances between colleagues in dyadic exchange make the application of each-others knowledge less untactful. Additionally, it reveals that theories on geographical proximity and knowledge flow (Boschma, 2005; Kleinbaum et al., 2013) persist during times of unprecedented advances in digital communication. During the study, teams were almost exclusively meeting through online conferencing tools, and besides the arduous time-differences present, a colleague normally at the next desk was as far away as a colleague on the other side of the world. The effects of geographical dispersion remain, at least in the short to medium term, a relevant contextual factor in the next era of digitization.

Additionally, range was operationalized similar to (Reagans & McEvily, 2003), but with a twist; since areas of expertise, and the interconnection between those areas, could not be established in this study, departmental network diversity was used, a measure slightly different than what has been used in literature before, and more tailored to the siloed context of the study. Surprisingly, in-betweenness centrality, a measure similar to the one used for range in (Tortoriello et al., 2012), turned out to not be a significant predictor, when it was coincidentally added as a check. This finding, combined with the unique use of the range variable in this study, outlines the value of contextual embeddedness of range, suggesting that in some organizational contexts a different measure of network diversity can be more appropriate. This study hypothesized range as an indicator of the ability of senders and receivers to integrate knowledge, which would decrease the need for cohesion, in line with (Reagans & McEvily, 2003). Instead, the heterogenuous context of a diverse network calls for cohesion, to succesfully integrate the knowledge and apply it to make impact. These findings are in accordance with the R-K index of (Tortoriello et al., 2015), a multiplication of the redundant elements (cohesion) and knowledge diversity (range), showing that range plays an important role in more heterogeneous contexts than the one studied in (Tortoriello et al., 2015).

Findings on tie strength were mostly in accordance with literature. This study anticipated an interaction effect between tie-strength and cohesion, but rejected this hypothesis. This is in line with (Levin et al., 2016), who only found a moderating effect between cohesion and trust, but not through closeness and frequency, and with (Reagans & McEvily, 2003; Tortoriello et al., 2012) that tie-strength and cohesion seem to mostly operate in parallel, each providing their unique contributions to dyadic knowledge flow.

5.2.2 Practical implications

This study provides practical insight in the way that interpersonal knowledge flows can be supported within cross-departmental interdependent groups within a large, heterogeneous bounded network of people interdependent on product knowledge.

Tacitness was by far the strongest and most robust predictor of the receipt of useful knowledge. This study has shown that the tacit and undocumented nature of knowledge inhibits effective flow in advise seeking networks within the usecase. Employees indicated that documentation often does not need its requirements, and there is a mismatch between the creation of a document and how it is used. Integrating alignment initiatives, that reflect on the knowledge needs of documents, and how to increase their productive use, in the continuous improvement portfolio will help employees share and find the knowledge they need to do their work effectively. Knowledge that cannot be codified will remain inhibited in its flow, and an awareness of the stickiness of this kind of knowledge calls for suitable time investment of proper transfer of crucial knowledge when it is tacit.

Despite the restricted use of office workplaces during the collection of data, due to the ongoing Covid-19 pandemic, geographical proximity turned out to be a strong predictor of the receipt of useful knowledge. Teams adapting online communication tools rapidly did not prevent timezone differences, cultural differences, local cohesiveness and team-membership from impacting knowledge flow. This study showed that geographical distance was the second most robust predictor of succesful knowledge flow.

In a cross-departmental, siloed context, such as the one described in the usecase, managers are recommended to foster the development of common understanding and language surrounding cross-departmental collaboration. From interviews and discussions with stakeholders, it became apparent that the siloed departmental context inhibits the flow of knowledge across-departments. In this study, several department related variables, such as departmental network diversity, as well as proximity, an presumed accumulation of communication difficulty (Reagans & McEvily, 2003), homophilic knowledge search (Kleinbaum et al., 2013) and cultural differences (Boschma, 2005), showed to be influential in the effective flow of knowledge. Tasselli et al. (2020) show in their recent study that, next to structural bridges between organizational boundaries, such as interpersonal ties or cohesion, the adoption of similar vocabularies between employees can help collaborate across soloed departments. This study showed the difficulty within the usecase company to establishing a definition of a cross-departmental network, and the challenge of establishing a cross-departmental projection of knowledge interdependence, due to these very vocabularies varying between groups and contexts. Data for the 'Areas of expertise' variable were unusable, because a Knowledge Managers agreed that employees' descriptions their competences diverged too much to create effective categories of common expertise. Establishing common vocabularies could aid in further breaking down the silos of NPD-networks.

Literature shows several ways in which this can be accomplished. One way of establishing common vocabularies is to align values across departments (Painter, Pouryousefi, Hibbert & Russon, 2019). Another is to give groups of individuals the authority and resources to break down barriers (Gulati, 2010), by attributing more decision power to business-lines, to create or support knowledge-broker groups, or by supporting a quality team to build a common quality standards across the departments. Lastly, putting the customers' needs central in all departments (Gulati, 2010), and by creating cross-departmental goals toward customer satisfaction that the departments can only meet collaboratively, can also help groups share knowledge more actively.

5.2.3 Recommendations for the usecase company

Recommendations for the case-company are split into two; the first set of recommendations are manegerial implications for all managers who find themselves working with similar NPD-based networks. These are general recommendations, but written toward the context of the usecase company. The second set of recommendations is based specifically on network- and contextual descriptive information that is sensitive to the specific context in which it is relevant. These recommendations can be found in appendix A.

The first practical recommendation, based on the data present at the case company, is that the dataset for this study could be used to calculate tie occurrence probabilities, using exponential random graph models. While out of the scope of this study, tie-occurrence analysis would be highly complementary to the current study as it shows what factors influence the initiation of ties in the first place. Appendix A presents recommendations based on qualitative evaluations of tie-occurrences between groups, but a quantitative assessment would be more thorough, valid, and contribute to literature by also assessing the unique use-case in a similar study to how (Aalbers, Dolfsma & Koppius, 2013) predict inter-unit ties and (Nakauchi et al., 2017) look at the difference between ties between and across departmental boundaries. In the continuation of this kind of studies, the dependent and independent variables can also be separated easier, since workplace analytics can be used to acquire data on tie occurrence, and survey-data can provide the contextual information on independent variables needed to test hypotheses beyond the studies of (Aalbers et al., 2013; Nakauchi et al., 2017). This avenue of research represents opportunities that both ask less (or more directed) time of employees in the gathering of survey data, and more complete and dynamic data on collaboration networks

Following this recommendation, the second big opportunity for practitioners of the usecase, but also for practitioners in similar usecases, is the continuation of measurement over time. The continuous improvementment group is currently addressing perceived documentation issues, as well as closing geographical boundaries. An investigation on the network resulting from these interventions has great evaluative power, and would, additionally, add to literature's strong call for longitudinal studies (Wijk et al., 2008; Phelps et al., 2012). While these studies can be more challenging (Stadtfeld, Snijders, Steglich & van Duijn, 2020), they can aid in methodological gaps in literature (Phelps et al., 2012).

5.2.4 Limitations

This thesis has several important limitations, that impact its validity and generalizability to other contexts. These limitations are described below.

Firstly, as outlined in section 3.5.3, the survey study conducted is subject to common method bias. Participants indicated that the survey was very long, and while 30 minutes was the indicated completion time, the researcher was even challenged by one of the teams to complete it in the allotted time; it was clear that the survey was perceived as an inconvenience, which could have let to miss-attributions or omissions. Given the limited sample size and the amount of variables in the models presented, this measurement error could potentially severely impact the reliability of results.

Secondly, as the cross-validated regression models showed in figure 4.3, the data is subject to severe model instability, another symptom of a small number of observations given the ambitious tests in this study. Inconsistencies in answers, as shown in the negative skew in the "less-successful" interactions figure 3.7b and indications by participants that the survey was long, might have been causes of this instability. It seems that the study, which has lost significant amount of data due to missingness, either in incomplete interactions, self-reference, unprovided or invalidly provided data, has a dataset which is small for the investigation it attempts to accomplish. Unfortunately, this diminishes generalizability of the results.

Thirdly, this study does not completely adequately control for the interdependence of observations present in the data. Both WLS models and MLM indicate the weight of raters' influence on regression estimates, they do not completely control for it. For the MLM, only rater-based groups were assigned a random effect. A more approproate way would have been to assign both sender and receiver a random effect, or a fixed effect, to completely control for individual deviations (Simpson, 2001). Unfortunately, for MLM to work properly, a majority of group-sizes of only one observation limits the effectiveness of MLM too much. An an initial test with fixed effect models for both sender and receiver showed signs of strong over-fitting. The present study did not have the data available to fit enough fixed or random effects for interdependence of observations to be properly addressed. This decreases the confidence in regression coefficient estimates, as the actual number of observations is arguably lower than that presented (Hair et al., 2010).

Fourthly, this study was unable to control for endogeneity, which puts important limitation on the interpretation of results, and an even larger limitation on the implication of causality (Hair et al., 2010). Therefore, omitted variable bias is a serious concern for coefficient estimates in the regression analysis. Several potentially important individual characteristics of dyadic knowledge flow have been omitted, and will be discussed accordingly; personality, and motivation. People do not always feel that sharing knowledge is in their interest, even if other conditions support it. Knowledge is a source of power in organizations, and sharing can be experienced as a weakening of personal value to a company, or employees might feel that they can avoid sharing by free-riding (Hislop, 2009). Employees must feel that there are incentives for the transfer of knowledge (Argote et al., 2003), especially when it concerns tacit knowledge (Osterloh & Frey, 2000), in a way that is proportional to their experienced cost of transfer (Ghosh & Rosenkopf, 2015), in an environment that is safe for personal risk-taking and speaking up (Edmondson, 1999). Literature on silos in organizations [citation needed] and discussions in workshops conducted within ASML showed that cross-departmental incentives were often not present. Incentives, therefore, could play a role in both the problem, as well as a solution, to knowledge flow challenges, of which the effect cannot be quantified from this study. This study measured variables that might be related to or result from motivation, and there are no strong indications from exploratory interviews or employees on motivational issues, there might still be endogenous factors at play whose effects cannot be measured in this study. Any follow-up within, or in similar contexts, is therefore advised to control for motivation.

Quite some control variables were added on respondents, to represent the unique characteristics that might influence results. There are however, more personal attributes, that could have influenced the results of the study, related to personality and individual capability. Reagans and McEvily (2003); Levin and Cross (2004) control for this problem by introducing a fixed effect for all respondents, and while this was the original plan for this study as well, the lack of data obstructed this powerful control of respondent-specific effects. As a result, factors like big-five attribute scores (Matzler, Renzl, Müller, Herting & Mooradian, 2008) and absorptive capacity (Cohen & Levinthal, 1990; Szulanski, 1996; Wijk et al., 2008), which are related to the successful transfer of knowledge, were omitted in this study.

To nuance the above, the characteristics of dyadic knowledge flow, as outlined in this thesis, are not as absolute as presented. For instance, while network diversity is presented as a structural characteristic of dyads, Reagans and McEvily (2003) argue that it represents personality ability; the skill to convey ideas to a more diverse set of colleagues. Analogously, Reagans and McEvily (2003) view social cohesion as a promoter of the motivation to invest time and effort in knowledge sharing. Tacitness, the way it is operationalized by Hansen (1999) and used in this study, also related to extent to which the alter is able to codify knowledge. Network-size is also related to personality, as personality traits predict the in-degree of peoples networks (Fang et al., 2015). Even still, while the result of motivation and personality might partially be represented by factors included in the study, the fact remains that these effects cannot be untangled, and there is no way to determine with certainty the endogeneity of these effect. This does not necessarily inhibit predictive validity of the model in its context, but severely limits causal conclusions, and the generalizability of specific variable findings to our overall-understanding of knowledge flow.

Literature has established the important role cognitive distance can have in the transfer of knowledge (Cohen & Levinthal, 1990; Boschma, 2005; Hislop, 2009) and network studies have consistently controlled for it, finding significant effects, for instance in role similarity (Tasselli et al., 2020; Tortoriello et al., 2012; Levin et al., 2016). The interviews conducted seemed to agree with the idea that the integration between perspectives is a difficult line to walk in this usecase. The variable of cognitive distance, representing both the extent to which the people in the dyad were part of the same department and functional background, did not predict the receipt of useful knowledge, however. This could have been the case due to issues with the data collected on cognitive distance. Measures like educational dissimilarity, field of expertise and product team membership did not result in usable data, and as a result, a part of cognitive distance between employees could not be measured. Especially product team membership, indicating whether recipients of knowledge were part of a team that centrally integrates knowledge for product development decisions, could have resulted in interesting findings.

Lastly, the response rate achieved in this study is on the low side of acceptability. Ideally, response rates are are at least 80%, and it could be argued that any diversion from a complete response of a bounded network could lead to important omissions in the network (Borgatti et al., 2006; Grannis, 2010). This study did evaluate the resulting network on differences in groups of respondents and non-respondents, but this does not reveal the impact of ommission completely.

5.2.5 Future research

This study identified a number of promising avenues of future research. Below, academic recommendations, relating to future avenues of research, are provided.

Firstly, this study established the gap in literature associated with the limited understanding of how complexity interacts with relational and structural characteristics of dyadic knowledge flow beyond tie strength (Hansen, 1999; Phelps et al., 2012). In general, knowledge characteristics are the least investigated characteristics in knowledge transfer literature (Phelps et al., 2012). Causal ambiguity, or the uncertainty of how components of knowledge interact, has been the most prominent knowledge characteristic studied (Wijk et al., 2008; Konlechner & Ambrosini, 2019), but this literature still calls for an understanding of the optimal levels of tacitness, complexity and specificity for the successful transfer of knowledge (Konlechner & Ambrosini, 2019). While this study was unsuccessful in establishing a general itemset to measure complexity, which was admittedly ambitious given the scope of this project, its investigation reveals the tenacity of complexity in its embeddedness in context, and the divergence in its definition and measurement in past literature. The establishing of heterogeneity, interdependence and uniqueness as its important components hopefully aids future researchers in integrating complexity in their dyadic knowledge flow studies, and to uncover how complexity interacts with other knowledge characteristics to reduce ambiguity (Konlechner & Ambrosini, 2019) and improves interpersonal transfer (Phelps et al., 2012).

While this study presented the evaluation of knowledge flow characteristics in an NPD-network, something new to literature, a review of social network analysis and organizational innovation projects found that network studies including external parties, on an individual level (i.e. team members with customers), is very limited (Leenders & Dolfsma, 2016). A second opportunity for future research lies in use-cases where an NPD-network is an open, collaborative effort with external parties in tightly involved ecosystems or open innovation projects, would greatly add to the findings in this study. It is expected that further barriers to recombination, as well as a stronger conflict between collaboration and competition (Ritala & Hurmelinna-Laukkanen, 2013), will significantly influence how contextual factors influence dyadic knowledge flow and knowledge diffusion across project teams from separate entities.

This study calls for the continuation of the investigation into friction and redundance in dyadic knowledge flow. This study showed that the relationship between cohesion and the separate stages of dyadic knowledge flow has been established pretty well, and generally, as the stages move from transfer, acquisition and application, it seems that cohesion turns from positively to nullified or even negatively related to dyadic knowledge flow. Three things remain unclear however. The role of cohesion in the first stage of dyadic knowledge flow, the search for knowledge, remains unclear. Ties identified in the survey studies were ties that were already established, and what made employees know and value others' knowledge enough to engage in tie formation (Borgatti & Foster, 2003a) remains unclear. Cohesion might, similarly to distance and organizational boundaries (Kleinbaum et al., 2013), predict tie formation over time, impacting dyadic knowledge flow.

Secondly, this study based the hypothesis of an inverted U-shape relationship on the separate effects of cohesion on friction and redundance. While quite some studies have been able to show the relationship between cohesion and knowledge flow outcomes, less emphasis is placed on what the relationship between friction and redundance separately looks like for different stages in dyadic knowledge flow. These effects will be hard to untangle, but would deepen our current understanding of how cohesion brings about dyadic knowledge flow.

Thirdly, there is an inherent limitation in the focus of studies on the use of subjective measures in the estimation of friction and redundance. Likert-scale evaluations are embedded in context, and reflect knowledge flow, but also reflect the reviewers' expectation in relation to what was achieved. It might be the case that redundance is less observable, and that decisions that were taken that should not have been made, and the knowledge that was not sought or integrated that would have made the project more of a success, tell a story in more objective measures of dyadic knowledge flow. Regrettably, the objective measurement of dyadic knowledge flow seems to be hard to accomplish, as current literature measures subjective measures from the perspective of the receiver (Tortoriello et al., 2015), sender (Reagans & McEvily, 2003), or a third observer (Hansen & Løvås, 2004), or objective measures that can only evaluate diffusion, not flow, because they are not directly related to an interaction, and only imply it (Singh, 2005; Sorenson et al., 2006). An alternative method where the evaluation is more controlled, in laboratory experiments for instance, like (Fahrenkopf, Guo & Argote, 2020), could help researchers reach objectivity while still measuring the interaction itself, not its implied occurrence (Sorenson et al., 2006) or its result.

Lastly, opportunities of alternative network edge data collection exist, that might deepen understanding of cohesion. Both cohesion values and number of contacts identified in the survey study are comparable to Reagans and McEvily (2003). Even studies like that of (Reagans & McEvily, 2003) and other cohesion studies do not uncover all ties; employees know potentially hundreds of colleagues that they do not directly collaborate or communicate with frequently, but these peripheral network structures might still represent a significant effect on cohesion that remains endogenous in current studies. Multiplexity, or the identification of other ties between employees that run in parallel of knowledge seeking ties, could similarly complement current findings on network structure (Paruchuri, Goossen & Phelps, 2019).

The present study found the managerial role to significantly predict the receipt of useful knowledge, in accordance with (Tortoriello et al., 2012). The simplistic nature of the measure in this study, which combined all team leads and project leads in one group, might have caused some nuance to be lost on how roles impact the receipt of useful knowledge. An interesting avenue for future research is to investigate the origin of the effect this has on the receipt of useful knowledge. It could be that managers represent generalists, and non-managers specialists, in this context. This could influence their access to superior knowledge or organization-specific knowledge (Fahrenkopf et al., 2020), which could in its place influences knowledge application.

References

Aalbers, R., Dolfsma, W. & Koppius, O. (2013). Individual connectedness in innovation networks: On the role of individual motivation. *Research Policy*, *42*(3), 624–634.

Alavi, M. & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107–136.

Alguezaui, S. & Filieri, R. (2010). Investigating the role of social capital in innovation: sparse versus dense network. *Journal of knowledge management*.

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

Argote, L. & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes*, *82*(1), 150–169.

Argote, L., McEvily, B. & Reagans, R. (2003). Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management science*, *49*(4), 571–582.

Argote, L. & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science*, *22*(5), 1123–1137. doi: 10.1287/orsc.1100.0621

Borgatti, S. P., Carley, K. M. & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social networks*, *28*(2), 124–136.

Borgatti, S. P., Everett, M. G. & Johnson, J. C. (2018). *Analyzing social networks.* Sage.

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

Borgatti, S. P. & Foster, P. C. (2003b). The network paradigm in organizational research: A review and typology. *Journal of management*, *29*(6), 991–1013.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, *39*(1), 61–74.

Burt, R. S. (2001). Structural holes versus network closure as social capital. In *Social capital* (pp. 31–56). Routledge.

Burt, R. S. (2004). Structural holes and good ideas. *American journal of sociology*, *110*(2), 349–399.

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization science*, *15*(5), 555–568.

Case-company. (2019). Research knowledge management program.

Case-company. (2021a). Case description - improving the productivity of knowledge flows.

Case-company. (2021b). Finding, sharing and maintaining knowledge.

Castaneda, D. I., Manrique, L. F. & Cuellar, S. (2018). Is organizational learning being absorbed by knowledge management? A systematic review. *Journal of Knowledge Management*, *22*(2), 299–325. doi: 10.1108/JKM-01-2017-0041

Cohen, W. M. & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128–152.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, *94*, S95–S120.

Cowan, R. & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of economic Dynamics and Control*, *28*(8), 1557–1575.

De Holan, P. M. & Phillips, N. (2004). Remembrance of things past? The dynamics of organizational forgetting. *Management Science*, *50*(11), 1603–1613. doi: 10.1287/mnsc.1040.0273

du Plessis, M. (2007). The role of knowledge management in innovation. *Journal of Knowledge Management*, *11*(4), 20–29. doi: 10.1108/13673270710762684

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350–383.

Fahrenkopf, E., Guo, J. & Argote, L. (2020). Personnel mobility and organizational performance: The effects of specialist vs. generalist experience and organizational work structure. *Organization Science*, *31*(6), 1601–1620.

Fang, R., Landis, B., Zhang, Z., Anderson, M. H., Shaw, J. D. & Kilduff, M. (2015). Integrating personality and social networks: A meta-analysis of personality, network position, and work outcomes in organizations. *Organization science*, *26*(4), 1243–1260.

Filieri, R. & Alguezaui, S. (2014). Structural social capital and innovation. is knowledge transfer the missing link? *Journal of knowledge management*.

Ganco, M. (2013). Cutting the gordian knot: The effect of knowledge complexity on employee mobility and entrepreneurship. *Strategic Management Journal*, *34*(6), 666–686.

Ganguly, A., Talukdar, A. & Chatterjee, D. (2019). Evaluating the role of social capital, tacit knowledge sharing, knowledge quality and reciprocity in determining innovation capability of an organization. *Journal of Knowledge Management*.

Ghosh, A. & Rosenkopf, L. (2015). Perspective—shrouded in structure: challenges and opportunities for a friction-based view of network research. *Organization Science*, *26*(2), 622–631.

Grannis, R. (2010). Six degrees of “who cares?”. *American Journal of Sociology*, *115*(4), 991–1017.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, *78*(6), 1360–1380.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic management journal*, *17*(S2), 109–122.

Gulati, R. (2010). *Reorganize for resilience: Putting customers at the center of your business.* Harvard Business Press.

Gupta, A. K. & Govindarajan, V. (2000). KNOWLEDGE FLOWS WITHIN MULTINATIONAL CORPORATIONS. , *496*(August 1999), 473–496.

Haans, R. F., Pieters, C. & He, Z.-L. (2016). Thinking about u: Theorizing and testing u-and inverted u-shaped relationships in strategy research. *Strategic management journal*, *37*(7), 1177–1195.

Hair, J. F., Anderson, R. E., Babin, B. J. & Black, W. C. (2010). *Multivariate data analysis: A global perspective (vol. 7).* Upper Saddle River, NJ: Pearson.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2019). Multivariate data analysis.

Haldin-Herrgard, T. (2000). Difficulties in diffusion of tacit knowledge in organizations. *Journal of Intellectual capital*.

Hansen, M. T. (1999). The Search-Transfer Problem : The Role of Weak Ties in Sharing Knowledge across Subunits Organization. , *44*, 82–111.

Hansen, M. T. & Løvås, B. (2004). How do multinational companies leverage technological competencies? moving from single to interdependent explanations. *Strategic Management Journal*, *25*(8-9), 801–822.

Hansen, M. T., Mors, M. L. & Løvås, B. (2005). Knowledge sharing in organizations: Multiple networks, multiple phases. *Academy of Management journal*, *48*(5), 776–793.

Hennig, M., Brandes, U., Pfeffer, J. & Mergel, I. (2012). *Studying social networks: A guide to empirical research.* Campus Verlag.

Hislop, D. (2009). *Knowledge management in organizations: A critical introduction.* Oxford university press.

Humphries, M. D. & Gurney, K. (2008). Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. *PloS one*, *3*(4), e0002051.

Jiménez-Jiménez, D. & Sanz-Valle, R. (2011). Innovation, organizational learning, and performance. *Journal of business research*, *64*(4), 408–417.

Kang, M. & Kim, Y.-G. (2010). A multilevel view on interpersonal knowledge transfer. *Journal of the American Society for Information Science and Technology*, *61*(3), 483–494.

Kao, T.-W. D., Su, H.-C. & Chen, Y.-S. (2019). The curvilinear relationships between structural embeddedness and productive efficiency: An exploratory study. *International Journal of Production Economics*, *212*, 176–185.

Kim, S. & Anand, J. (2018). Knowledge complexity and the performance of inter-unit knowledge replication structures. *Strategic Management Journal*, *39*(7), 1959–1989.

Kleinbaum, A. M., Stuart, T. E. & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316–1336.

Konlechner, S. & Ambrosini, V. (2019). Issues and trends in causal ambiguity research: a review and assessment. *Journal of Management*, *45*(6), 2352–2386.

Kreft, I. G. & De Leeuw, J. (1998). *Introducing multilevel modeling.* Sage.

Leenders, R. T. & Dolfsma, W. A. (2016). Social networks for innovation and new product development. *Journal of Product Innovation Management*, *33*(2), 123–131.

Leonard, D. & Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California Management Review*(3), 112–132. doi: 10.2307/41165946

Levin, D. Z. & Cross, R. (2004). The Strength of Weak Ties You Can Trust : The Mediating Role of Trust in Effective Knowledge Transfer .
doi: 10.1287/mnsc.1030.0136

Levin, D. Z., Walter, J., Appleyard, M. M. & Cross, R. (2016). Relational enhancement: How the relational dimension of social capital unlocks the value of network-bridging ties. *Group & Organization Management*, *41*(4), 415–457.

Lewis, K. & Herndon, B. (2011). Transactive memory systems: Current issues and future research directions. *Organization science*, *22*(5), 1254–1265.

Liebowitz, J. (1999). Key Ingredients to the Success of an Organization ' s Knowledge Management Strategy. , *6*(1), 37–40.

Marsden, P. V. & Campbell, K. E. (1984). Measuring tie strength. *Social forces*, *63*(2), 482–501.

Matzler, K., Renzl, B., Müller, J., Herting, S. & Mooradian, T. A. (2008). Personality traits and knowledge sharing. *Journal of economic psychology*, *29*(3), 301–313.

McEvily, B., Soda, G. & Tortoriello, M. (2014). More formally: Rediscovering the missing link between formal organization and informal social structure. *Academy of Management Annals*, *8*(1), 299–345.

McFadyen, M. A., Semadeni, M. & Cannella Jr, A. A. (2009). Value of strong ties to disconnected others: Examining knowledge creation in biomedicine. *Organization science*, *20*(3), 552–564.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.

Mors, M. L. (2010). Innovation in a global consulting firm: When the problem is too much diversity. *Strategic Management Journal*, *31*(8), 841–872.

Nakauchi, M., Washburn, M. & Klein, K. (2017). Differences between inter-and intra-group dynamics in knowledge transfer processes. *Management Decision*.

Nonaka, I. & Peltokorpi, V. (2006). Objectivity and subjectivity in knowledge management: a review of 20 top articles. *Knowledge and process management*, *13*(2), 73–82.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford university press.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V. & Van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research policy*, *36*(7), 1016–1034.

Osterloh, M. & Frey, B. S. (2000). Motivation, knowledge transfer, and organizational forms. *Organization science*, *11*(5), 538–550.

Painter, M., Pouryousefi, S., Hibbert, S. & Russon, J.-A. (2019). Sharing vocabularies: Towards horizontal alignment of values-driven business functions. *Journal of Business Ethics*, *155*(4), 965–979.

Paruchuri, S., Goossen, M. C. & Phelps, C. (2019). Conceptual foundations of multilevel social networks.

Perry-Smith, J. E. (2006). Social yet creative: The role of social relationships in facilitating individual creativity. *Academy of Management journal*, *49*(1), 85–101.

Phelps, C., Heidl, R. & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of management*, *38*(4), 1115–1166.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Powell, W. W., Staw, B. & Cummings, L. (1990). Neither market nor hierarchy.

Reagans, R. & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly*, *48*(2), 240–267.

Ritala, P. & Hurmelinna-Laukkanen, P. (2013). Incremental and radical innovation in coopetition—the role of absorptive capacity and appropriability. *Journal of Product Innovation Management*, *30*(1), 154–169.

Rivkin, J. W. (2001). Reproducing knowledge: Replication without imitation at moderate complexity. *Organization Science*, *12*(3), 274–293.

Schulz, M. (2001). The uncertain relevance of newness: Organizational learning and knowledge flows. *Academy of management journal*, *44*(4), 661–681.

Simpson, W. (2001). Qap: The quadratic assignment procedure. In *North american stata users' group meeting* (pp. 12–13).

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, *51*(5), 756–770.

Sorenson, O., Rivkin, J. W. & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. doi: 10.1016/j.respol.2006.05.002

Stadtfeld, C., Snijders, T. A., Steglich, C. & van Duijn, M. (2020). Statistical power in longitudinal network studies. *Sociological Methods & Research*, *49*(4), 1103–1132.

Swan, J., Newell, S. & Scarbrough, H. (1999). Knowledge management and innovation : networks and networking. , *3*(4), 262–275.

Swart, J. & Kinnie, N. (2003). Sharing knowledge in knowledge-intensive firms. *Human resource management journal*, *13*(2), 60–75.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal*, *17*(S2), 27–43.

Tasselli, S. & Caimo, A. (2019). Does it take three to dance the tango? organizational design, triadic structures and boundary spanning across subunits. *Social Networks*, *59*, 10–22.

Tasselli, S., Zappa, P. & Lomi, A. (2020). Bridging cultural holes in organizations: the dynamic structure of social networks and organizational vocabularies within and across subunits. *Organization Science*, *31*(5), 1292–1312.

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H. & Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain connectivity*, *1*(5), 367–375.

Tortoriello, M. & Krackhardt, D. (2010). Activating cross-boundary knowledge: The role of simmelian ties in the generation of innovations. *Academy of Management Journal*, *53*(1), 167–181.

Tortoriello, M., McEvily, B. & Krackhardt, D. (2015). Being a catalyst of innovation: The role of knowledge diversity and network closure. *Organization Science*, *26*(2), 423–438.

Tortoriello, M., Reagans, R. & McEvily, B. (2012). Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units. *Organization science*, *23*(4), 1024–1039.

Tsoukas, H. (1996). The firm as a distributed knowledge system: A constructionist approach. *Strategic management journal*, *17*(S2), 11–25.

Tushman, M. L. (1977). Special boundary roles in the innovation process. *Administrative science quarterly*, 587–605.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American sociological review*, 674–698.

Wang, J.-C., Chiang, C.-h. & Lin, S.-W. (2010). Network structure of innovation: can brokerage or closure predict patent quality? *Scientometrics*, *84*(3), 735–748.

Wang, S. & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human resource management review*, *20*(2), 115–131.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440–442.

Wei, J., Zheng, W. & Zhang, M. (2011). Social capital and knowledge transfer: A multi-level analysis. *Human Relations*, *64*(11), 1401–1423.

Wijk, R. V., Jansen, J. J. P. & Lyles, M. A. (2008). Inter- and Intra-Organizational Knowledge Transfer : A Meta-Analytic Review and Assessment of its Antecedents and Consequences. (June).

Yu, Y., Hao, J.-X., Dong, X.-Y. & Khalifa, M. (2013). A multilevel model for effects of social capital and knowledge sharing in knowledge-intensive work teams. *International Journal of Information Management*, *33*(5), 780–790.

Zander, U. & Kogut, B. (1995). Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test. *Organization Science*, *6*(1), 76–92. doi: 10.1287/orsc.6.1.76

Zhao, Z. J. & Anand, J. (2013). Beyond boundary spanners: The 'collective bridge'as an efficient interunit structure for transferring collective knowledge. *Strategic Management Journal*, *34*(13), 1513–1530.

A. Appendix - Confidential

This appendix includes all matters taken out of the master thesis because of confidentiality. This appendix includes an elaboration on the setting of the study, shows more network descriptives that are sensitive if the context is know, and discuss more practical recommendations for ASML based on more detailed network findings.

A.0.1 Introduction and problem relevance

ASML is the largest semiconductor company in the Netherlands, and a global leader in the development, manufacturing and distribution of lithography machines. Their products offer customers the capacity to produce highly advanced integrated circuits. The development of lithography machines on this level is highly complex, and high market demands for development speed forces ASML to go through rapid product improvement and steep learning curves to remain competitive (Case-company, 2021a).

Product development projects are initiated in, and go trough, the Product Generation Process (PGP). ASML has a broad range of departments, but four departments carry the responsibility of the operational work of product development. Development and Engineering (D&E), Manufacturing (MF), Sourcing & Supply Chain (S&SC) and Customer Service (CS)[1], are the departments over which key decision gates in the PGP move. These departments are often called 'the sectors', and together with the customer as end-point, they form the value chain of ASML. A new development initiative in the PGP housed in one of the three major business lines; EUV, DUV or Applications (Apps). Business-lines, who are the product and project owners of PGP projects moving along the value chain, are profit and loss accountable for the projects in the PGP.

ASML knowledge management argues that the segmentation of the sectors has shown to structurally limit the efficient flow of knowledge across functional and departmental boundaries. Connecting knowledge, created in the firm, with where it is needed, is becoming increasingly more difficult. The incomplete and untimely availability of knowledge across-sectors impacts the efficiency individual employees, as they spend up to 18% of their time searching for the right knowledge, and introduces feelings of frustration (Case-company, 2019). As a recent study in the employee experience of finding and maintaining knowledge has shown, rapid growth in employee numbers, together with high time pressure, proliferate these issues (Case-company, 2021b). While ASML sets industry standards in its ability to innovate, knowledge inefficiencies accumulate on the level of company goals and business stakeholders agree that customer satisfaction with delivered products is still a frontier where improvement is possible. Issues in quality, serviceability or manufacturability still often occur during development, and their occurrence, especially in later development stages, or worse yet, at the customer site, has significant business impact.

The challenges experienced by ASML knowledge management resonate with the field of knowledge management literature, which argues that organizational knowledge has become one of the key assets of businesses (Hislop, 2009). Especially for knowledge-intensive firms like ASML, (Swart & Kinnie, 2003), organizational knowledge is deemed one of its key assets (Nonaka & Takeuchi, 1995; Leonard & Sensiper, 1998). The knowledge based view of the firm poses that the effective integration of individuals' knowledge in goods and services is the primary role of organizations

[1] This is a simplification, but for this study it is valid enough

(Grant, 1996), and given that organizations are dynamic, distributed knowledge systems, then "getting connected and interrelating the knowledge each [person] has [...] the key to achieving coordinated action" (Tsoukas, 1996, p.22).

This effective redistribution and application of knowledge is exactly what is strained by organizational silos and barriers, however. As Kleinbaum et al. (2013) showed, in an unprecedented review of digital communication data between employees, organizational boundaries of units and functions have a significant effect on communication frequency. Functional and departmental boundaries often result in this soli-ing effect, where communication is strained (Allen et al., 2007; McEvily et al., 2014). Especially in innovation, novelty introduces ambiguities that make interpretation across boundaries more arduous (Carlile, 2004). It is especially at these boundaries where fostering knowledge flow can be of crucial importance, as Carlile (2004) concludes that "innovation occurs at the boundaries between specialized domains [and] effectively managing knowledge across the various types of boundaries in an organization is what drives competitive advantage"(Carlile, 2004, p.566). Accordingly, the importance of understanding how knowledge flow is established grows as departmental, functional and geographical boundaries become more prominent, and the potential inhibition of the flow of knowledge gets more severe.

A.1 The usecase

One of the smaller businesslines within ASML is home to many of the company-wide challenges in efficient and effective knowledge flow across organizational silos. Within this smaller business-line, especially fast growth and a decreasing relative number of senior engineers, amplified knowledge management challenges. As a consequence, ASML knowledge management has decided to investigate knowledge management challenges for business-line stakeholders.

Business-line managers are convinced that these issues need to be addressed. Together with people from continuous improvement and quality, they specified a system module, a component of a machine that ASML produces, as a pilot group. It is the ambition of business-line managers to scale findings from this usecase to the other system modules in the business-line. The the usecase is representative of many cross-departmental product development chains within ASML. The group, employing around 250 people across all sectors, presents a highly heterogenous subject group. With three large work-sites, in Europe, The United States and Asia, it covers the three largest work-sites for the businessline. Additionally, the group represents a mix of hard-ware and software, with a roughly 50/50 split in D&E working on hardware and complementary software.

The cross-sectoral usecase consists twenty-three [2] interdependent teams with the collaborative interdependence of developing, manufacturing, supplying and servicing the optical column, part of several product models and development projects. It is important to note that these groups were implicitly, but not explicitly, defined as being part of the same system module; this was a D&E definition, *projected on other sectors*, based on interdependence.

[2] At least twenty-three teams. It was surprisingly hard to define exactly which groups contributed, because the system function is a D&E specific definition. Other sectors are organized differently. In other sectors, there might be teams that contribute to the entire product, with only one or two people contributing to the system module. Because of this ambiguity, S&SC, for instance, is counted as one team in this definition, even though from an S&SC perspective, that is not correct.

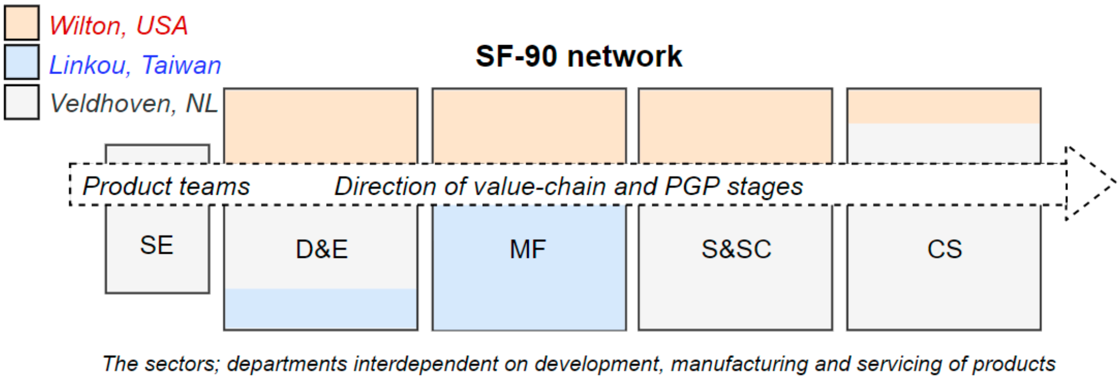


Figure A.1: The NPD-network of the usecase
A matrix organization of the sectors, interdependent departments, and the product teams, that coordinate collaboration and are profit and loss accountable on the projects in the network. Colours are not fully representative of employee numbers, for a precise network boundary, see 4.4.1

Figure A.1 shows how the sectors, and their corresponding teams, are centered around the PGP, the stage-gate new product development (NPD) process. Sector representatives acknowledge the importance of timely availability of the right knowledge within this process. In this view, the effective creation and utilization of organizational knowledge is one of the key drivers of organizational performance, essential for innovation (Jiménez-Jiménez & Sanz-Valle, 2011). The PGP forms the spinal-cord of the knowledge interdependence between the teams. Key decision moments in the PGP are the places where all knowledge related to the optical column, its past performance, and how it should be improved to enhance customer satisfaction, should come together. At the start of a new product development project, which is run by specified teams within D&E, requirements are needed from other sectors, so as to prevent issues down the development pipeline, and prevent mistakes made earlier. As the development project matures, another important interdependence is the handover of D&E to MF, S&SC and CS. The latter sectors need to be properly informed on the system specifications and inner-workings, to be able to do their work well. The effective application of knowledge is crucial in product innovation processes like these (Cohen & Levinthal, 1990; du Plessis, 2007; Jiménez-Jiménez & Sanz-Valle, 2011).

These knowledge needs are serviced in the PGP with formal knowledge sharing structures. Key decisions at the end of each stage in the PGP process have a list of formalized documentation as input and output, which should contain all knowledge needed for requirement integration and effective handovers. This documentation is supplemented with formal meeting structures, with the aim of integrating the required perspectives for key decision making. While the formal PGP process is a thoroughly matured stage-gate process, it is not, and cannot be, perfect. Stakeholders indicate that handovers are often incomplete or incomprehensible. Documentation is limited or too detailed, and continuous sense-making and perspective taking is required to complement these documents. They also argue that that the right people are not involved from the start, or that people with complementary perspectives in other sectors are unknown. As a result, arguably important input is not taken into account at the right moments, or misunderstood, and rejected without consideration. To complement the inevitable shortcomings of formal structures, employees naturally engage in interpersonal discussion and iteration. In an instance in manufacturing where representatives indicated alignment and sharing was going well, they elaborated that at least three cross-sectoral alignment meetings a week were required to facilitate the perspective-taking and understanding required to work together.

Busisness-line management and supporting staff have currently initiated an investigation in formal and managerial exchange processes within the usecase. Formal structures do not represent the totality of organizational structures that facilitate knowledge exchange, however. As Allen et al. (2007) put it, “formal channels of communication rarely accurately reflect the working relationships between individuals [...], the myriad of personal communications and ties which in reality disseminate knowledge and information between individuals constitute the informal social networks” (Allen et al., 2007, p.180). McEvily et al. (2014) contribute to this point, adding that “the intricate interplay between formal and informal elements is [...] what ultimately determines individuals’ ability to get things done and, consequently, their capacity to facilitate (or, at times, hinder) the pursuit of organizational objectives”. (McEvily et al., 2014, p.333). Informal relations form a unique conduit over which knowledge flows, as “informal ties promote vicarious learning. Informal connections allow people to benefit from knowledge accumulated by close contacts” (Argote et al., 2003, p.576).

The role of organizational collaboration networks in the diffusion of information and knowledge has been one of the central themes in social network theory since its inception (Powell et al., 1990; Borgatti & Foster, 2003a). Theories on network closure (Coleman, 1988; Burt, 2004) brokerage (Burt, 2004), boundary spanning (Tushman, 1977), embeddedness (Uzzi, 1996) social capital (Coleman, 1988) all conceive interpersonal network structures to be important in the access to and flow of information resources. Knowledge management literature has acknowledged the importance of relational aspects of knowledge for knowledge sharing, and recognizes that the active context of learning takes place in networks of members, tasks and tools (Argote & Ingram, 2000; Argote & Miron-Spektor, 2011).

The important role of social networks is recognized by employees as well, as ASML is often called a ‘network organization’. In the use-case, this interaction is driven by informal knowledge seeking and sharing to elaborate on the content, or to supply information that is not present, or cannot be found, or cannot be written down in digital knowledge repositories. In their day-to-day work, engineers of all sectors also need each-others’ expertise to resolve issues as they come up.

While revision of formal structures is important, without a good view on informal structures, no holistic set or interventions can be designed, and informal blind-spots might cause omission of relevant organizational aspects, or over-treatment of a knowledge gap in formal structure that social networks already effectively close. As a result, ASML Knowledge Management, who is supporting the initiative, has argued that the workshop on formal and managerial knowledge should be complemented by a review of informal knowledge seeking interactions.

A.2 Informal knowledge flows within the usecase

While stakeholders of the system module recognize that these sharing networks can be very effective at diffusing knowledge across-sectors, there are several contextual challenges that impede the flow of knowledge. Literature recognizes that, while networks themselves are structures of interpersonal interactions (Borgatti & Foster, 2003b), contextual factors have an effect how how effectively they diffuse knowledge (Szulanski, 1996; Ghosh & Rosenkopf, 2015). One of these characteristics is relational, and concerns in what ways which two people engaged in knowledge exchange relate to eachother (Szulanski, 1996; Levin & Cross, 2004; McPherson et al., 2001; Boschma, 2005; Reagans & McEvily, 2003). Another one is structural, and describes how the personal network is embedded within the complete network of interactions (Reagans & McEvily, 2003; Zhao & Anand, 2013; Kim & Anand, 2018). The last characteristic is the knowledge being diffused, and to what extend it is hard to share and interpret by others (Hansen, 1999; Zander & Kogut, 1995; Sorenson et al., 2006). An investigation into the informal knowledge flow characteristics of the usecase revealed that relational, structural and knowledge characteristic factors played a role in how effective and efficient knowledge could flow between people, across team-, sector- and location boundaries.

Relational aspects mostly related to distance, both cognitive as well as physical. As a representative sourcing and supply chain indicated, geographic barriers made transfer and alignment cost more time. "Out of sight, out of mind" - is a way an interviewee of D&E Wilton described this difficulty. A representative of S&SC remarked that "when you spread teams across different parts of the world, with time difference, different cultures, it has to be managed. [...] it requires more alignment". ", the further away, the larger the problem of course", a representative of D&E Veldhoven remarked about the collaboration with Asia. Proximity literature has shown that these effects can influence the effective flow of knowledge, arguing that close proximity exposes actors to positive externalities, and facilitates informal relationships (Boschma, 2005).

A representative of S&SC indicated that common knowledge is needed to understand each-other. "Often", a D&E representative elaborates, "complexity is not even relevant, because you do not even get to that level". As a result, "80% of our time is spent understanding each-other", says a SE representative, responsible for integration cross-sectoral product requirements. The degree to which people are able to understand each-others perspective through common perspectives and knowledge, also termed cognitive distance, has been identified as an important relational factor when it comes to the effective flow of knowledge. Theory on cognitive distance shows that common knowledge is required to be able to integrate knowledge (Cohen & Levinthal, 1990), especially in such an interdisciplinary network.

Knowledge characteristics were also deemed an influential factor by interviewees. ASML is home to a high concentration of very specialized and highly educated people, to work on advanced, technologically complex systems (Case-company, 2021a). A recent study on the employee experience of knowledge management found that complexity was one of the major barriers to effective knowledge seeking and sharing (Case-company, 2021b). A representative of D&E Wilton remarked that "You cannot completely hide complexity", and an interviewee of manufacturing in Wilton stated that complexity can be an issue when sharing takes place between people with a difference in educational background. Complexity has indeed been shown to play a role in the successful flow of knowledge, especially between organizational units (Hansen, 1999; Hansen & Løvås, 2004), like in the use-case.

While the role of structural characteristics were less evident from interviews, having 'the right network' was something interviewees stressed was important. A D&E manager, who recently moved to the business-line, underlined the necessity of knowing the right people, indicating that he was dependent on others to connect him to knowledge he needed. In contrast, the inability to connect the "right people" to meetings and discussions was seen as a barrier to improving decision quality.

It is widely known that these contextual factors play an important role in the diffusion of knowledge. Not enough is known, however, *to what extent* the contextual factors of the usecase impact the ability of employees within the optical column usecase to exchange knowledge effectively. A clear image of how these contextual factors obstruct and influence the effectiveness of knowledge flows, and the successful utilization of cross-sectoral knowledge between sectors, remains elusive.

The remainder of this chapter remains the same, and can be found in section 1.3 in the first chapter of the main report.

A.2.1 The bounded network

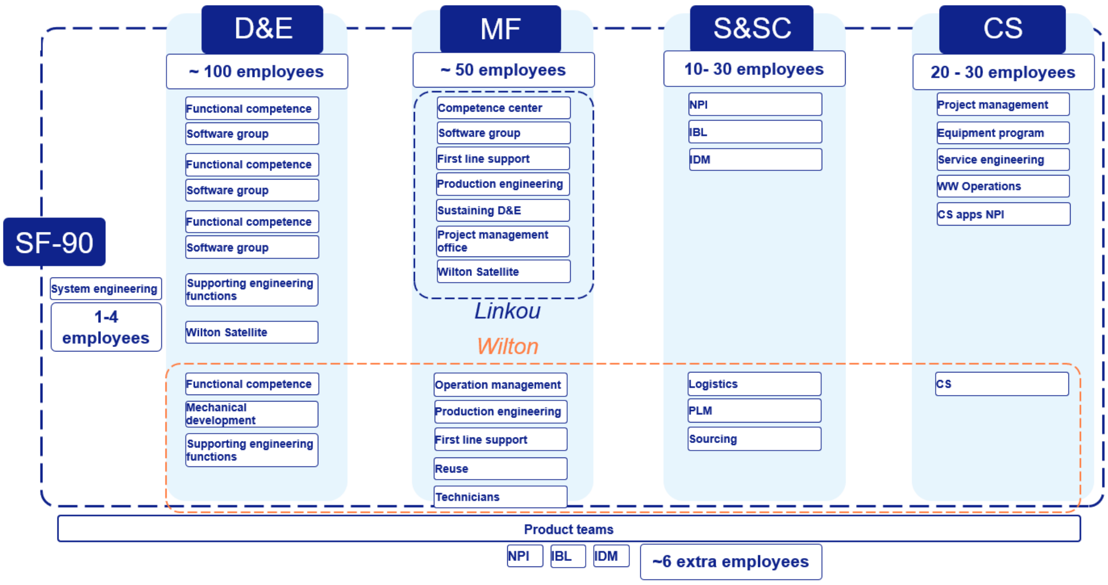


Figure A.2: The bounded network groups included in the study
Groups not marked with a location were situated in The Netherlands

A.2.2 Network visualizations

In general, the following findings are drawn from network visualizations:

- The network is centralized around D&E. Access to knowledge goes through central connectors.
- Most groups are not directly connected to each-other, but access knowledge through other groups. This puts emphasis on the need of ties to conduct the flow of knowledge as, otherwise, knowledge doesn't diffuse to all groups equally (and equally fast)
- Access to inforation varies widely between groups and individuals. To what extent this is a problem for those individuals depends on their role, needs and their relation to the bounded network.

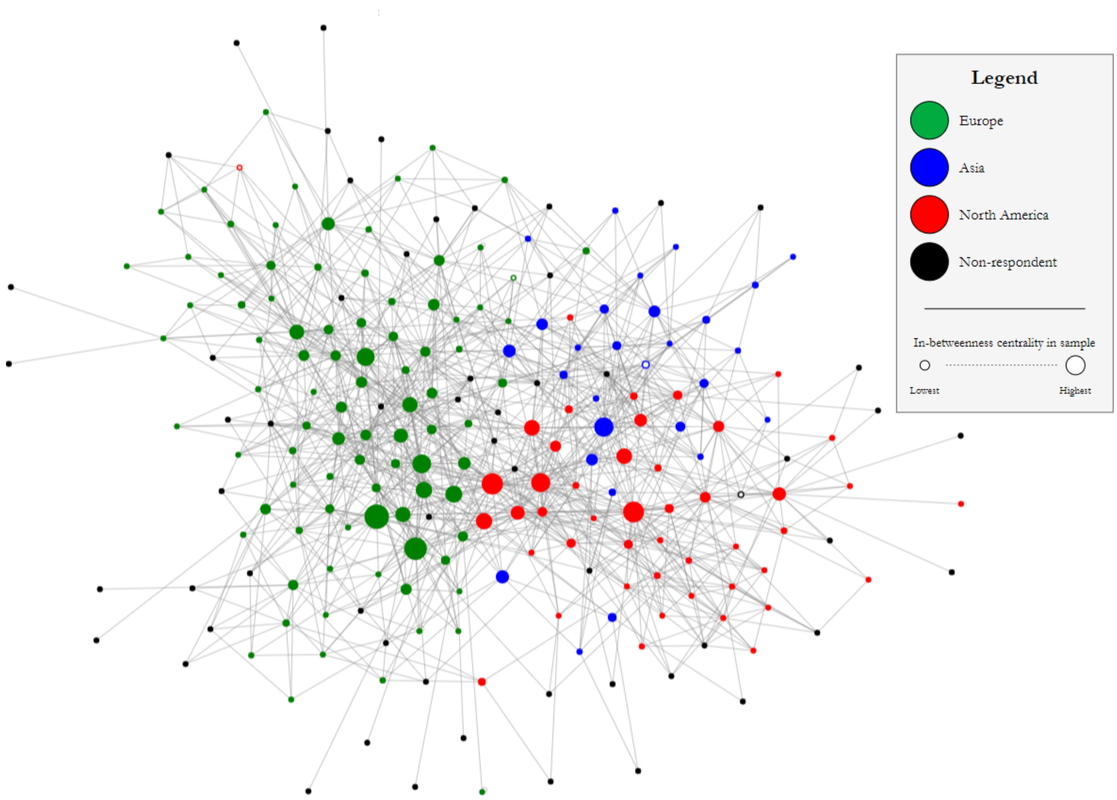


Figure A.3: The unbounded location network

The following findings are drawn from the location visualization:

- There is a clear divide in the network between the Netherlands on one side, and the USA and Asia on the other. Heatmaps show that Asia and the US are relatively better interconnected to each-other than to the Netherlands.
- nodes become central in the network when their network is spread across multiple locations. This suggests locational brokers play a key role in knowledge diffusion.

The following findings are drawn from the sectoral visualization:

- MF and S&SC are more peripheral in the network than D&E, CS and the product teams.
- Wihtin the same sector, there can be large differences in informational access.

The following findings are drawn from the team visualization:

- Within D&E Veldhoven, Software has much less influence in the network than their corresponding functional clusters do. Between Asia and the US, the difference seems to be less pronounced.
- Industrial engineering and the satellite team are more centrally connected in the clusters where they are supposed to be, which is positive.
- Fuctional cluster groups seeem to be most influential, however, at least in terms of being the

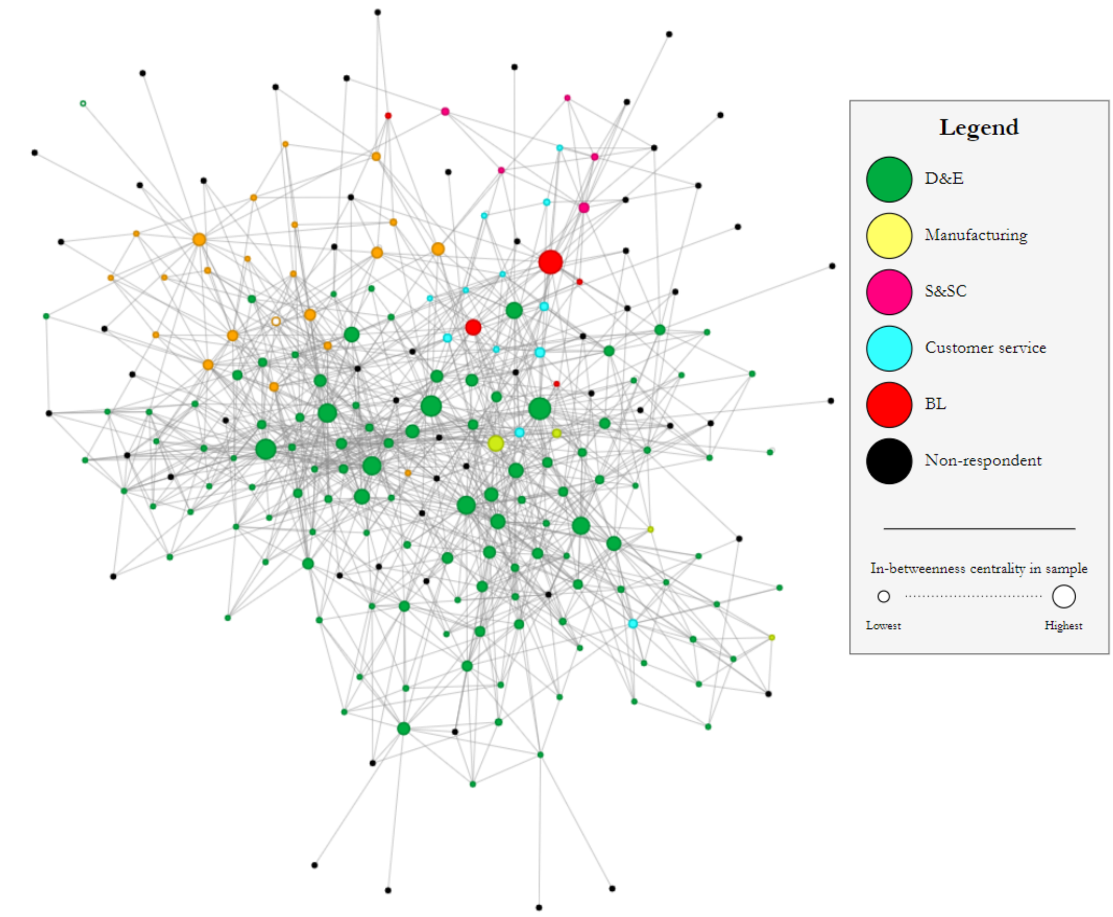


Figure A.4: The unbounded departmental network

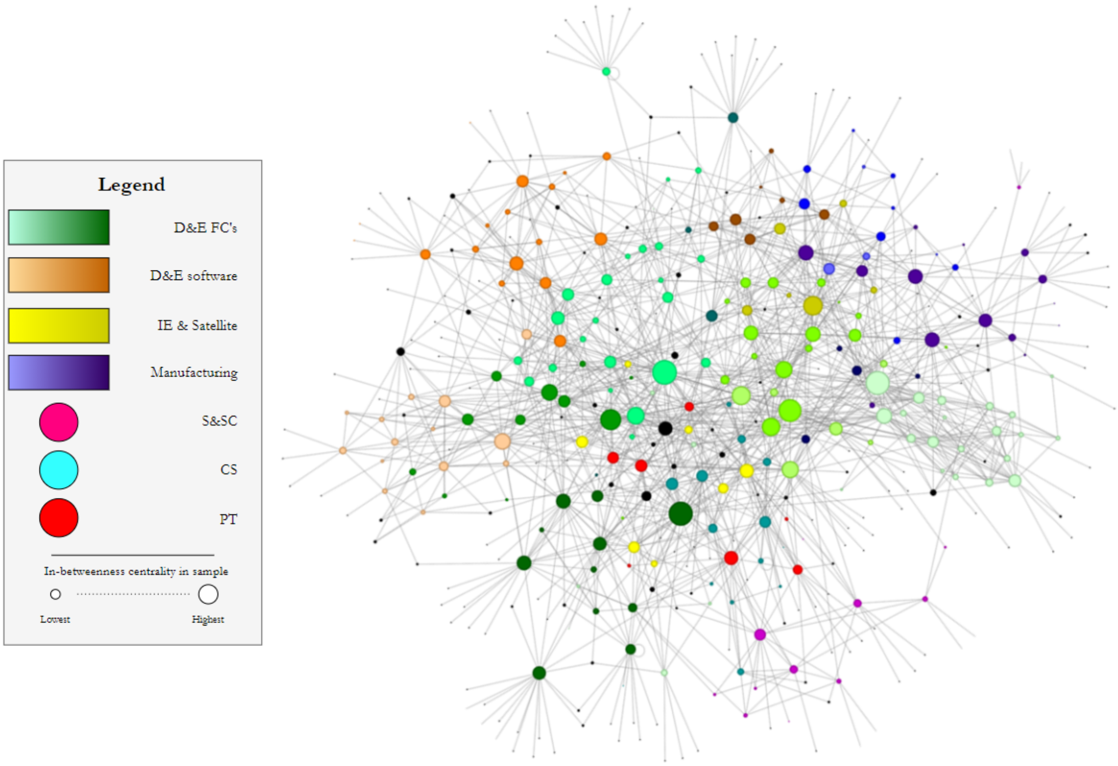
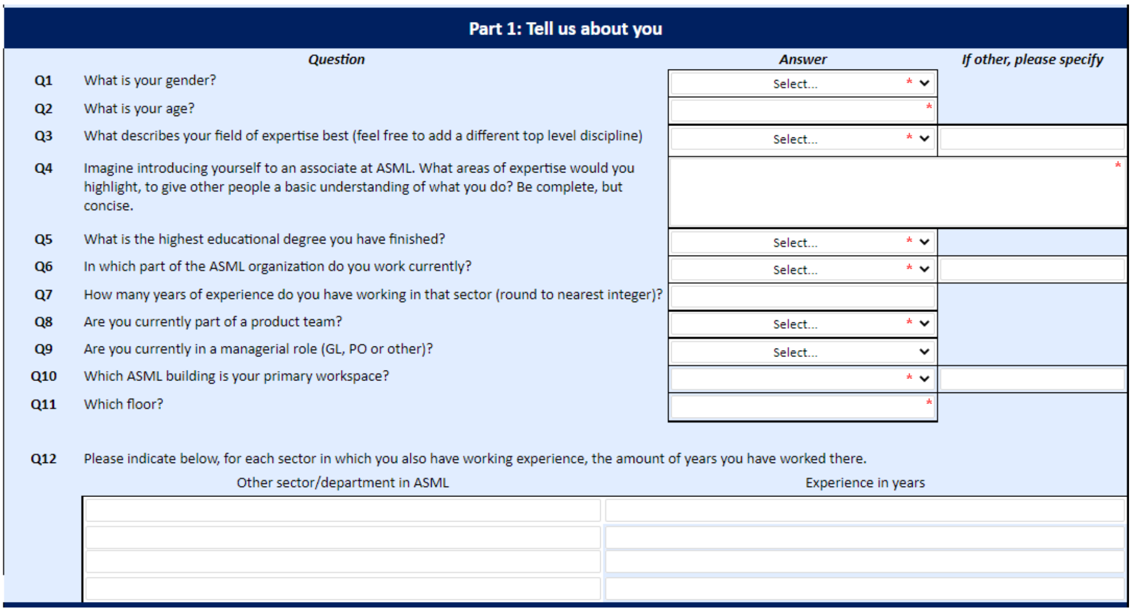


Figure A.5: The unbounded team-network

sources of knowledge.

B. Appendix - Full survey

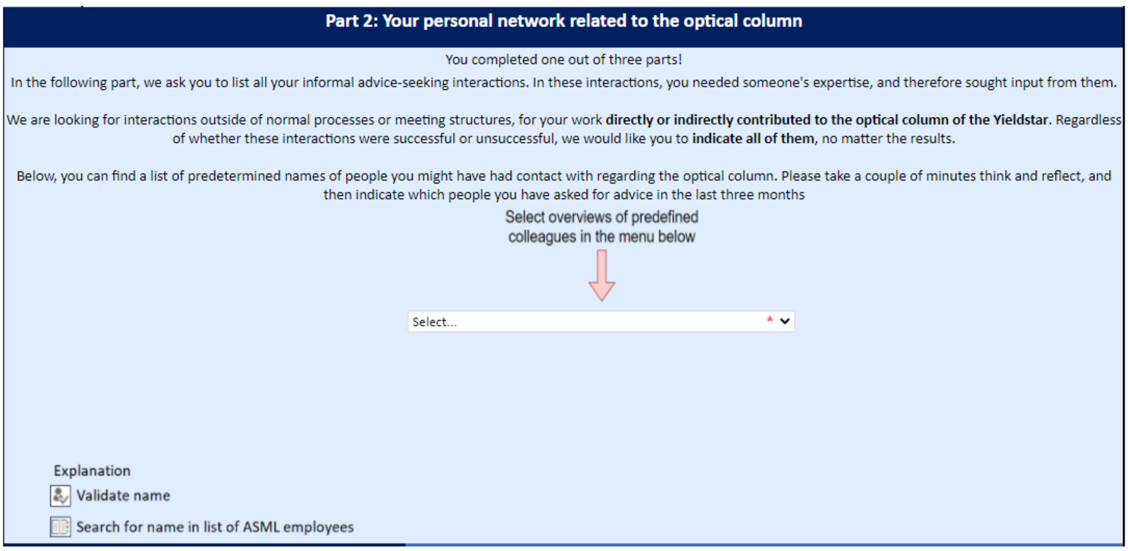


Part 1: Tell us about you

	Question	*Answer*	*If other, please specify*
Q1	What is your gender?	Select... *	
Q2	What is your age?	*	
Q3	What describes your field of expertise best (feel free to add a different top level discipline)	Select... *	
Q4	Imagine introducing yourself to an associate at ASML. What areas of expertise would you highlight, to give other people a basic understanding of what you do? Be complete, but concise.	*	
Q5	What is the highest educational degree you have finished?	Select... *	
Q6	In which part of the ASML organization do you work currently?	Select... *	
Q7	How many years of experience do you have working in that sector (round to nearest integer)?		
Q8	Are you currently part of a product team?	Select... *	
Q9	Are you currently in a managerial role (GL, PO or other)?	Select...	
Q10	Which ASML building is your primary workspace?	*	
Q11	Which floor?	*	

Q12 Please indicate below, for each sector in which you also have working experience, the amount of years you have worked there.

Other sector/department in ASML	Experience in years



Part 2: Your personal network related to the optical column

You completed one out of three parts!

In the following part, we ask you to list all your informal advice-seeking interactions. In these interactions, you needed someone's expertise, and therefore sought input from them.

We are looking for interactions outside of normal processes or meeting structures, for your work **directly or indirectly contributed to the optical column of the Yieldstar**. Regardless of whether these interactions were successful or unsuccessful, we would like you to **indicate all of them**, no matter the results.

Below, you can find a list of predetermined names of people you might have had contact with regarding the optical column. Please take a couple of minutes think and reflect, and then indicate which people you have asked for advice in the last three months

Colleague you asked for advise in the last three months	How often do you interact with this person?				How close do you feel to this person?			
	Daily	Weekly	Monthly	Less often/did not interact before	Especially close	Close	Less than close	Distant
*								
*								
*								
*								

Are there any important people that are not on the list of pre-defined names? Feel free to add them in the list below.

Colleague you asked for advise in the last three months	How often do you interact with this person?				How close do you feel to this person?			
	Daily	Weekly	Monthly	Less often/did not interact before	Especially close	Close	Less than close	Distant

Interaction 1: one of the successful interactions

Q1 Describe in one sentence what you sought advise on

Q2 With whom did you have this interaction? (this is someone you mentioned in part 2)

Q3 Is this someone with whom you get together for informal social activities outside work? Select...

Q4 The information I received from this colleague made (or is likely to make) the following contribution to (choose one per statement)

	Contributed very negatively	Contributed negatively	Contributed somewhat negatively	Contributed neither positively nor negatively	Contributed somewhat positively	Contributed positively	Contributed very positively
(internal) client satisfaction							
Project team performance							
The project's value to my organization							
The project's quality							
The project coming in on budget or closer to coming in on budget							
Reducing costs in the project							
My ability to spend less time on the project							
Shortening the time this project took							

Q5	Indicate to what extend you agree with the following statements about the knowledge you sought from this person				
	Agree	Somewhat agree	Don't agree or disagree	Somewhat disagree	Disagree
You do not have to be a domain expert to be able to understand and apply this knowledge	○	○	○	○	○
A lot of prerequisite knowledge, or a large list of instructions, is needed to effectively understand and apply this knowledge	○	○	○	○	○
This knowledge is specific for the domain in which I work	○	○	○	○	○
The work that I needed advice on requires a lot of time spent on considering how different aspects or components of the work interaction and influence each-other	○	○	○	○	○
The work I needed advice on is <u>not highly</u> dependent on different disciplines, interests or perspectives	○	○	○	○	○
This knowledge <u>does not</u> depend on a large amount of different areas of expertise to be understood and applied	○	○	○	○	○

Q6 How well documented was the knowledge you received from this person?

0 - It was very well documented			3 - it was somewhat well documented			6 - It was not well documented
○	○	○	○	○	○	○

Q7 Was all this knowledge sufficiently explained to you in writing?

0 - All of it was			3 - some of it was			6 - None of it was
○	○	○	○	○	○	○

Q8 What type of knowledge did you receive from this person?

0 - Mainly reports, documents, manuals			3 - Half know-how and half documents			6 - Mainly personal practical know-how
○	○	○	○	○	○	○