

Simensen, Erlend Osland; Abbasiharofteh, Milad

**Article — Published Version**

## Sectoral patterns of collaborative tie formation: Investigating geographic, cognitive, and technological dimensions

Industrial and Corporate Change

**Provided in Cooperation with:**

Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale)

*Suggested Citation:* Simensen, Erlend Osland; Abbasiharofteh, Milad (2022) : Sectoral patterns of collaborative tie formation: Investigating geographic, cognitive, and technological dimensions, Industrial and Corporate Change, ISSN 1464-3650, Oxford University Press, Oxford, Vol. 31, Iss. 5, pp. 1223-1258,  
<https://doi.org/10.1093/icc/dtac021> ,  
<https://academic.oup.com/icc/article/31/5/1223/6593918>

This Version is available at:

<https://hdl.handle.net/10419/264601>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# Sectoral patterns of collaborative tie formation: investigating geographic, cognitive, and technological dimensions

Erlend Osland Simensen<sup>1,\*</sup> and Milad Abbasiharofteh<sup>2,3,4</sup>

<sup>1</sup>Centre for Technology, Innovation and Culture, University of Oslo, P.O. box 1108, Blindern, Oslo 0317, Norway. e-mail: [e.o.simensen@tik.uio.no](mailto:e.o.simensen@tik.uio.no), <sup>2</sup>Department of Economic Geography, Faculty of Spatial Sciences, University of Groningen, Landleven 1, Groningen 9747 AD, The Netherlands, <sup>3</sup>Department of Human Geography and Spatial Planning, Utrecht University, Princetonlaan 8a, Utrecht 3584 CB, The Netherlands and <sup>4</sup>Department Structural Change, Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser-Straße 2, Halle (Saale) 06120, Germany. e-mail: [m.abbasiharofteh@rug.nl](mailto:m.abbasiharofteh@rug.nl)

\*Main author for correspondence.

## Abstract

The literature on sectoral systems of innovation evidences the crucial importance of sectoral specificities in innovation. Also, the proximity framework enables the exploration and analysis of how various actors establish collaborative ties. However, there has not yet been much cross-fertilization between the proximity literature and sectoral systems of innovation. First, we conceptually integrate the proximity literature into sectoral studies by underlining the differences and commonalities across sectors. Second, we use a dataset of research and development (R&D) projects subsidized by the Research Council of Norway (RCN) between 2005 and 2016 to analyze the impact of geographical and cognitive proximities on collaborative tie formation in four Norwegian sectors (biotech, oil and gas, marine, and maritime). The results indicate that cognitive proximity and colocation in large cities are positively related to the creation of collaborative ties across all sectors. Yet, the joint effects of cognitive proximity and colocation in large cities vary across sectors and seem to be correlated with the complexity of sector-specific technologies.

**JEL classification:** O31, O38, D85

## 1. Introduction

Innovation is understood as a combinatorial process, where novelty emerges by combining existing knowledge and materials (Schumpeter, 1911; Weitzman, 1998). For innovation to happen there is an increasing need to collaborate with other firms and public research organizations because such collaborative relations enable firms to keep up with the ever-increasing complexity of technologies and the combinatorial requirements of innovation (Manniche *et al.*, 2017). Therefore, unraveling the complexity of knowledge networks, as a means of combining or recombining knowledge and materials, is one of the main challenges in understanding the contemporary economy (Bathelt and Glückler, 2018). For instance, empirical studies show that the position of an actor in a knowledge network correlates with its innovative performance (Breschi and Lissoni, 2009; Cantner *et al.*, 2010; Lazer and Friedman, 2016).

One of the key foundations of innovation theory is that the prerequisite for innovation is highly context-dependent and innovation scholars thus use the concept of innovation systems to

study the contexts in which innovations occur. *Sectors* are one such contextual boundary, and the literature on sectoral innovation systems has made a significant contribution to the understanding of how sectors vary in terms of innovation activities and the patterns of technological change (Breschi *et al.*, 2000; Malerba, 2002, 2005). It highlights the great heterogeneity of sectors and industries with respect to their innovation processes and knowledge sourcing<sup>1</sup> patterns. However, most empirical studies on collaborations and collective learning do not address the multiplicity of sectors in one cohesive empirical setting (e.g., Giuliani, 2011; Balland *et al.*, 2015a; Lazzeretti and Capone, 2016; Capone and Lazzeretti, 2018; Giuliani *et al.*, 2018). Given various context-specific factors in these studies, it is a challenging task to identify sector-specific factors that foster collaborations. We argue that there is a rather untapped potential in using collaboration knowledge networks as a representation of sectoral systems. This is an empirical source that covers all the components of a sectoral system.

While the proximity literature (Boschma, 2005; Torre and Rallet, 2005) provides a novel conceptual framework to investigate the drivers of knowledge sourcing, it has not been used to provide a better understanding of commonalities and differences across sectors—most of the empirical findings are based on studies from one region or sector. We therefore set out to use the proximity approach and to empirically answer the following questions. How and to what extent do proximity dimensions influence the establishment of collaborative ties across sectors? How does the interplay between proximity dimensions (complementarity and substitutability effects) affect knowledge sourcing across sectors?

We investigated an extensive and unique dataset of research and development (R&D) projects subsidized by the Research Council of Norway (RCN)<sup>2</sup> between 2005 and 2016. The database includes various fields, such as the name and type of collaborators, the scientific domains of collaborations, and the duration of each project. Elaborating on this database, we applied the gravity model to analyze the impact of proximity dimensions on the formation of collaborative ties within four sectors: (i) biotech, (ii) oil and gas, (iii) marine, and (iv) maritime. The relation between geographical and cognitive proximities and the formation of collaborative ties is of the main interest in this study, while we operationalized and included institutional proximity as a control.

Our results show that the cognitive proximity is positively correlated with the creation of collaborative ties across sectors, whereas the geographical proximity seems not to be relevant in most cases. Instead, colocation in large cities is positively related to tie formation. More interestingly, the joint effects of the cognitive proximity and colocation in large cities vary across sectors and correspond to the complexity of sector-specific technologies.

This paper is structured as follows. Section 2 provides a discussion on the relevance of sectoral innovation systems and presents an overview of the proximity framework. Section 3 describes the empirical approach and the R&D activities in Norway. Section 4 presents the results and discusses the findings relevant to the proximity dimensions in the four sectors in Norway. Section 5 concludes the paper and discusses some limitations of this study, future research areas, and potential policy implications.

## 2. Sectoral systems of innovation and proximity dimensions

The sectoral approach has, compared to national, regional, and technological innovation systems, received less theoretical attention. This approach however is central in innovation studies and underlines the substantial role of sectoral actors, their interdependencies, and their collective cognitive boundaries as the drivers of change in a given sector (Breschi *et al.*, 2000; Malerba, 2002). Schumpeter stresses that firms are subject to sectoral characteristics in their innovation patterns (Schumpeter, 1911). More recently, Malerba (2002) defines sectoral change as a co-evolutionary process: “the elements of a sectoral system are closely connected, it follows that their change over time results in a co-evolutionary process of its various elements” (p. 259). This implies that the evolutionary development of sectors is partly determined by shifts in their

1 The process of searching for required knowledge and expertise by individuals and organizations.

2 In Norwegian: *Norges forskningsråd*.

cognitive boundaries, through interorganizational collaborations. [Malerba \(2005\)](#) and [Malerba and Adams \(2013\)](#) further argue that the main agents of sectoral change are (i) actors and networks, (ii) knowledge and technologies, and (iii) institutions. The attributes of these agents and the interplay between them determine a particular development trajectory for a sector. While this framework introduces the building blocks of sectoral systems, it does not account for how the complex process of knowledge sourcing (e.g., creating collaborative ties) functions—a process that lies at the heart of a sector's innovation processes and thereby contributes greatly to the future evolution of these systems.

[Pavitt \(1984\)](#) is a pioneering scholar whose taxonomy is based on the innovative behavior of firms, namely supplier-dominated, scale-intensive, specialized suppliers, and science-based firms. Later, one more group is introduced in his taxonomy to represent information-intensive firms (e.g., banking sector). While this taxonomy provides an analytical tool to understand how innovative behaviors and, to some extent, knowledge sourcing occurs across sectors, there are several issues that need to be addressed to better understand differences and commonalities across sectors regarding knowledge sourcing.

The first problem concerns the way in which the taxonomy is created. Pavitt observes innovative behavior patterns at the firm level and groups them at the industrial level, engendering a high degree of variance within each group ([Archibugi, 2006](#)). The second problem originates from the Pavitt's classification premise that firms have stable cognitive boundaries. Yet, through knowledge sourcing, firms evolve and might move from one taxonomy group to another because knowledge has a cumulative nature and potentially changes the cognitive boundary of a given firm ([Malerba, 2002](#)). The third problem concerns multi-technology or multi-products firms that can potentially be categorized in more than one group ([Archibugi, 2006](#)). For instance, Siemens is a well-known German company active in various fields ranging from telecommunications and energy to healthcare and biotechnology. The domain of activities of such firms goes beyond the sectoral boundaries and Pavitt's taxonomy.

Recent studies have aimed at alleviating these issues by widening the sectoral system approach to include users and institutions in the framework ([Geels, 2004](#)). Also, scholars have shifted their focus on sector-specific technological regimes, knowledge bases, cumulativeness, and appropriability conditions ([Breschi et al., 2000](#); [Asheim and Coenen, 2005](#); [Dosi et al., 2006](#); [Castellacci, 2008](#)). For instance, [Castellacci \(2008\)](#) builds on the concept of technological paradigm and “industry-specific regimes, trajectories, and vertical linkages” to provide a sectoral taxonomy that is more flexible than Pavitt's. Castellacci emphasizes the relevance of knowledge sourcing (“interactions” in his language) in industries close to “new technological paradigm” compared to the ones dominated by “less dynamic trajectories.” However, he says nothing about how these interactions differ across sectors. [Coad \(2019\)](#) empirically showed that firms from the same industrial sector do not converge toward the same R&D intensities. This finding resonates with the definition of sectoral systems as a place of diverse set of actors with relations crossing sectoral boundaries that serve different purposes for the industry. There are companies that specialize in delivering R&D for the industry, whereas other companies fulfill other roles and do not invest in R&D to the same extent. A newly published paper by [Rosiello and Maleki \(2021\)](#) combines the sectoral system approach with knowledge base complexity, similar to the ambition of the paper at hand. They find that technological catch-up varies with regard to the technological complexity of sectors, here they use patent data to classify technologies and sectors. Looking at the body of literature on sectoral systems, the authors of this article have failed to find attempts to use network data of knowledge actors for sectoral analyses, and we have found none of which compares several sectors in the same analysis.

This gap calls for an extension of the sectoral innovation framework that also includes the key driving forces of knowledge sourcing. This would stimulate an increased understanding of sectoral commonalities and differences. We argue in this paper that this extension can be built based on the proximity approach, which provides a strong conceptual framework for studying collaborative activities. It also has the capability of capturing the multi-layered interdependency among the various elements of sectoral systems of innovation.

The proximity literature provides a conceptual framework to better investigate the process of knowledge sourcing. [Boschma \(2005\)](#) theoretically argues that colocation is “neither a necessary

nor a sufficient condition for learning to take place” (p. 62). He also discusses how five proximity dimensions influence knowledge sourcing:

- Geographical proximity reflects how close the locations of two given organizations are or alternatively, whether or not they are located in the same geographical area (see [Micek, 2018](#) for a review).
- Cognitive proximity reflects how similar actors interpret and use new knowledge, based on their interpretation schemes and absorptive capacities ([Cohen and Levinthal, 1990](#); [Nooteboom, 2000](#); [Boschma, 2005](#)).
- Institutional proximity reflects the commonalities and differences at the macro level dominated by norms, values, recognized practices, and the rule of law ([Edquist and Johnson, 1997](#); [Boschma, 2005](#)).
- Organizational proximity is concerned with the extent to which organizations are similar in terms of autonomy and control in their organizational arrangements ([Boschma, 2005](#)).
- Social proximity refers to the extent to which organizations have shared experiences and are embedded in the same social networks, in which they benefit from lower transaction costs and increased synergy effects, trust, and altruistic behaviors ([Granovetter, 1985](#); [Coleman, 1988](#); [Uzzi, 1997](#)).

In addition to the above-mentioned five proximity dimensions, researchers defined alternative proximity dimensions based on the specificity of their theoretical arguments or empirical settings. These dimensions include virtual proximity ([Morgan, 2004](#)), technological proximity ([Greunz, 2003](#); [Usai et al., 2016](#)), socio-cultural proximity ([Gill and Butler, 2003](#); [Abbasiharofteh and Broekel, 2020](#)), and proximity on the move ([Bernela et al., 2019](#)). These alternative dimensions however are limited to one or few studies, whereas numerous empirical studies focus on the role of geographic and cognitive proximities and their relationship with the formation of collaborative ties (for a review, see [Balland et al., 2020a](#)).

Although the proximity dimensions are forces of different nature and analytically orthogonal, they may turn out to be positively correlated in practice ([Hardeman et al., 2015](#); [Balland et al., 2020a](#)). Thus, this is of critical importance to take into consideration the impact of different proximity dimensions to ensure that the strong explanatory power of one proximity dimension (e.g., geographical proximity) is not driven by other lacking correlated dimensions. Also, the proximity framework provides a conceptual engine to understand the interplay between different proximity dimensions ([Hardeman et al., 2015](#)). For instance, one could investigate how the likelihood of collaborative tie formation in a specific sectoral setting may increase (decrease) by the being proximate in more than one dimension. In other words, two proximity dimensions complement one another if being proximate in two dimensions increases the likelihood of tie formation (i.e., positive joint effects) and substitute one another if being proximate in the given proximities decreases the likelihood of tie formation (i.e., negative joint effects) ([Broekel, 2015](#); [Balland et al., 2015b](#); [Abbasiharofteh, 2020](#)). Anecdotally, two managers meet by chance because of working in the same geographic area, and they exchange knowledge about their projects. As a result, this face-to-face interaction helps them to bridge cognitive gaps ([Ter Wal and Boschma, 2011](#)). While this has been a long-standing wisdom, empirical studies of the joint effects of the proximity dimensions in various sectoral systems have been woefully few in number.

While many empirical studies have implemented this framework to comprehend the complexity of collaborative networks (for an overview, see [Balland et al., 2020a](#)). The proximity framework still does not provide a theoretically informed explanation for conflicting empirical results. That is, why one proximity dimension is a crucial factor in one sector and is not relevant or even negatively associated with knowledge sourcing in the other. For instance, [Balland et al. \(2013\)](#) empirically show that the relevance of geographical proximity persists in the video game industry. Conversely, in the case of the biotechnology sector, [Ter Wal's \(2014\)](#) study provides evidence that the geographical proximity loses its importance to social proximity. One might conjecture that the varying effects of proximity dimensions could emerge from various knowledge types and the attributes of dominant technologies in each sector. However, a more systematic

investigation of the commonalities and differences of sectors regarding the effect of proximity dimensions is still lacking (Boschma and Martin, 2010; Boschma, 2018).

Although systematic studies that investigate and compare the role of proximity dimensions in multiple sectoral setting is still lacking, the plethora of empirical studies investigating the role of geographical and cognitive proximity dimensions enable us to develop multiple lines of arguments on the varying role of geographical and cognitive proximities in different sectoral settings.

The relevance of geographical proximity for learning and knowledge diffusion has its origin in Marshall's (1920) work that underlines the importance of geographic collocation. Inspired by Marshall, a large body of literature shows that geographical proximity facilitates the process of tie formation (Micek, 2018). In the case of the Toy Valley cluster in Spain, Balland *et al.* (2015a) investigate the driving forces of tie formation in technical and business networks. They empirically show that geographical proximity facilitates knowledge sourcing performance in both networks. Molina-Morales *et al.* (2015) and Belso-Martínez *et al.* (2017) provide similar results in a foodstuffs cluster in the Valencian region. Juhász and Lengyel (2018) also observe the positive impact of geographical proximity in the printing and paper product cluster in Hungary. In the case of a high-tech cluster in Tuscany, Capone and Lazzaretto (2018) and Lazzaretto and Capone (2016) demonstrate that the impact of proximity dimensions on knowledge sourcing changes over time, with an increasingly positive effect of geographical proximity. Among the few studies at the sector level, Balland *et al.* (2013) demonstrate the positive impact of geographical proximity in the case of the global video game industry.

That said, one should not incorrectly conflate the effects of geographical proximity and agglomeration spillovers. Building on several empirical studies, Balland *et al.* (2020b) argue that empirical studies overestimate the effect of geographical proximity unless they control for the effects of other proximity dimensions (e.g., social proximity). Scholars argue that the relevance of geographical proximity is due to the fact that this proximity dimension facilitates mutual learning, by giving rise to social interaction and trust building (Jaffe *et al.*, 1993; Singh, 2005; Giuliani, 2011, 2013; Giuliani *et al.*, 2018).

The above argument is very much in line with the scholarly debate on the regional innovation system (Asheim *et al.*, 2011), which looks beyond benefits emerging from geographic collocation and takes into consideration a systematic view of knowledge sourcing, learning, and labor mobility networks embedded in agglomerations. Granovetter (1985) argues that such activities are mostly embedded in social networks which are geographically bounded. Thus, this spatial dimension of trust reduces transaction costs and eases involving in joint projects. Audretsch and Feldman (1996) provide empirical evidence for this argument by showing that innovative activities more strongly cluster compared to manufacturing employment. Thus, Micek (2018) argues that the concept of geographical proximity goes beyond the mere geographical distance between actors and should be seen in relation to the embeddedness of firms and individuals in a common spatial context, which fosters mutual learning. This line of argument leads us to the following hypotheses:

*Hypothesis 1: Both geographical proximity and collocation in large cities are positively related to collaborative tie formation between organizations.*

*Hypothesis 2: Collocation in large cities is more strongly related to collaborative tie formation than geographical proximity.*

Nooteboom (2000) describes how the extent to which organizations are cognitively proximate influences interorganizational learning. He argues that "information is useless if it is not new, but it is also useless if it is so new that it cannot be understood" (p. 72). This implies that the benefit of interorganizational learning is maximized only if optimum cognitive proximity is given (Wuyts *et al.*, 2005). To the best of our knowledge, the optimal cognitive proximity has not been addressed in the context of sectoral systems of innovation. In the context of Mode 2 knowledge, scholars argue that the process of knowledge production has increasingly become interdisciplinary. For instance, Gibbons *et al.* (1994) show that the average number of authors



per paper and their diversity regarding scientific, social, and geographical backgrounds have increased.

More recent studies on patent collaborations support the findings of Gibbons and his colleagues. Broekel (2019) provides several stylized facts about patent co-inventorship that point toward an increase in the relevance of cognitive proximity. That is, the complexity of patents increases over time, and this calls for more R&D and larger joint projects. For instance, van der Wouden's (2020) study addresses the history of co-inventorship in the United States between 1836 and 1975. This study shows that collaboration has increased in patenting since the mid-20th century, and there is a positive correlation between collaboration and the complexity of the filed patents. Perhaps the increase in the complexity and diversity of collaborations across all sectors accounts for the fact that the average technological distance between patents in metropolitan areas tends to decrease and that specialized cities have a denser collaboration network (van der Wouden and Rigby, 2019). While cognitively proximate organizations could benefit from inter-regional collaborations (e.g., Powell *et al.*, 1996), having a certain degree of cognitive proximity increasingly gains importance for creating collaborative ties. Thus, we suggest the following hypothesis:

*Hypothesis 3: Cognitive proximity is more strongly related to collaborative tie formation than geographical proximity and colocation in large cities.*

There is a scholarly debate on whether proximity dimensions could complement or substitute one another (Balland *et al.*, 2015b). Empirically, Ponds *et al.* (2007) show that one proximity dimension can compensate for the lack of another dimension in creating collaborative ties in a co-authorship network. Similarly, van der Wouden and Rigby (2019) provide evidence in the context of large urban agglomerations in the United States that the effect of social proximity substitutes the one of spatial proximity, whereas this substitutability is not given in smaller cities. Since the process of knowledge accumulation and diffusion depends on the attributes of underlying technologies in each sector (Dosi and Nelson, 2010), one could argue that the joint impact of cognitive proximity and the benefits of colocation in large cities on knowledge sourcing might be determined by sector-specific technologies and their complexity.

Balland and Rigby's (2016) study on patenting behavior in the United States between 1975 and 2010 gives evidence that supports this theoretical argument. That is, there is a direct link between the geographical distribution of activities and the complexity of underlying technologies because more complex activities agglomerate in large cities and metropolitan areas. This is plausible that the benefits of colocation in large cities facilitate the creation of collaborative ties in sectors dominated by complex technologies, which require larger teams consisting of experts from various related scientific fields (Balland *et al.*, 2015b). Yet, this is not the case in sectors dominated by less complex technologies, in which knowledge can be codified in different forms and diffused among cognitively proximate organizations over distance (Sorenson *et al.*, 2006). Based on the above discussion, we propose the following hypotheses:

*Hypothesis 4: The joint effects of colocation in large cities and cognitive proximity are positively related to collaborative tie formation in sectors dominated by more complex technologies.*

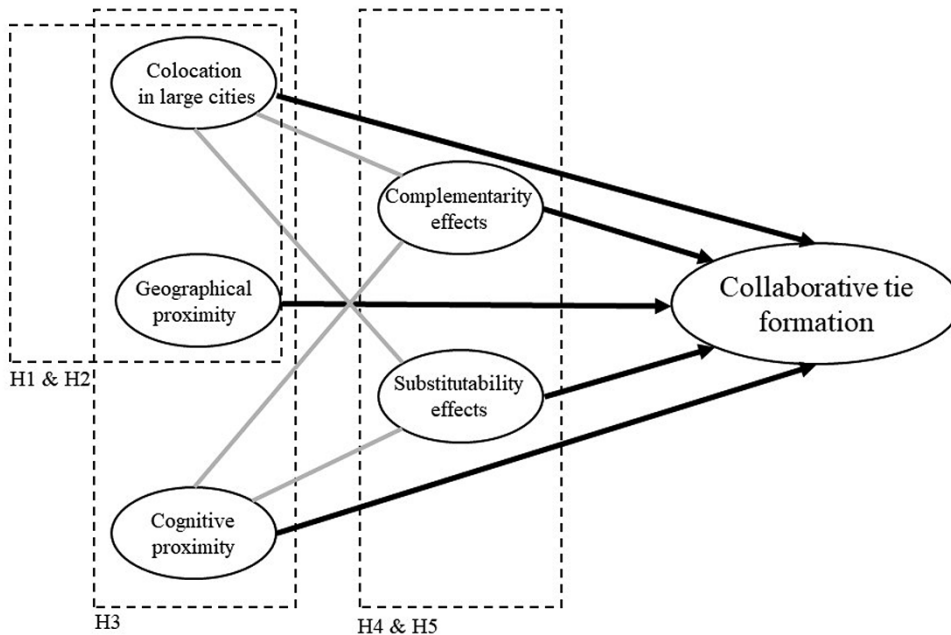
*Hypothesis 5: The joint effects of colocation in large cities and cognitive proximity are negatively related to collaborative tie formation in sectors dominated by less complex technologies.*

Figure 1 provides the nomological network that summarizes the hypotheses.

### 3. Empirical setting

#### 3.1 Data

In this paper, we investigate a collaboration network in the Norwegian innovation system. Norwegian research funding is distributed through two channels, namely basic research funding



**Figure 1.** Nomological network with corresponding hypothesis

for research institutions and the NCR.<sup>3</sup> This paper used data on R&D projects subsidized by the RCN between 2005 and 2016. This database includes 35,663 projects, of which 23,183 were carried out by more than one organization (i.e., joint projects). We assume that participants in a joint project create bilateral knowledge ties and mutually learn from one another, whereby we map knowledge exchange between various actors. These projects were conducted by 4623 actors,<sup>4</sup> ranging from private firms to publicly funded organizations.<sup>5</sup>

This database also provides one or several tags for each project showing the domain of activities in one or several fields. The RCN allocates multiple tags to every funded project, based on project descriptions and the domain of knowledge of the research project. There are three levels of tags (“topic group,” “topic,” and “subtopic”), and the lowest level includes over 125 tags. Many of these subtopics are sector-specific. We categorized four sectors at the subtopic level and used this categorization to determine to what sectors each project belongs. It is important to note that the definition of sectors in the data does not necessarily refer to the sector definition from finance or economics where they sort sectors from primary to tertiary. Our reading of Malerba’s definition is to a greater extent tied to knowledge bases of large industrial ecosystems. Delineations such as the standard European nomenclature of productive economic activities (NACE codes)<sup>6</sup> do not refer to a sectoral system’s dynamic and heterogeneous network of actors, nor do they refer to the knowledge bases and technologies that are the base of the sectoral system.

The share of private firms (80%) and the one of publicly funded organizations (20%) are roughly the same across all sectors. Since a project can span several sectors, it allows us to observe

3 The European Union framework program is an additional research scheme which is of much smaller magnitude, compared to the two main research programs in Norway.

4 After data cleaning, 4% of the actors could not be identified and were removed from the database.

5 We categorized actors in 13 categories: (i) association and industry organizations, (ii) banks, (iii) private firms and companies, (iv) county-level administrations, (v) government agencies, (vi) independent research institutes, (vii) municipalities, (viii) NGOs, (ix) public hospitals and health organizations, (x) public-private cluster organizations, (xi) schools and museums, (xii) university and university colleges (“høyskole”), and (xiii) others.

6 A common problem with the existing studies is thus the use of static indicators as the definition of sectors. For example, the widespread use of NACE codes or other static definitions of industrial sectors in empirical studies of sectoral systems are not very fitting for the concept. NACE codes are usually not reviewed over time and are part of a registration process as a company is founded. Some of the NACE codes are also generic (e.g., *Activities of head offices*).



**Table 1.** Number of R&D collaborative ties

	Biotech	Oil and gas	Marine	Maritime
Biotech	9520			
Oil and gas	143	8134		
Marine	482	563	13,546	
Maritime	248	201	735	8889

whether organizations collaborate in more than one sector. This is of great importance because it allows us to move beyond the traditional view of firms as being isolated in one sector and capture sectoral interdependencies.

Table 1 shows the number of collaborative ties (each organization pair) within and across four sectors. As shown, sectors show a varying degree of overlap. The largest overlap is shown between the marine and the maritime sector, and the lowest between oil and gas and biotech. This attribute of the data plays a crucial role because we can assign project-related information (e.g., the sector-specific knowledge domains of projects) to organizations and use this to create a cognitive portfolio for each actor, taking sectoral interdependencies into account.

Figure 2 shows the position of actors, as well as the collaborative ties at the beginning of four time windows, each spanning 3 years. Most organizations are located in one of four large Norwegian cities (Oslo, Bergen, Trondheim, and Stavanger).

### 3.2 Sectors

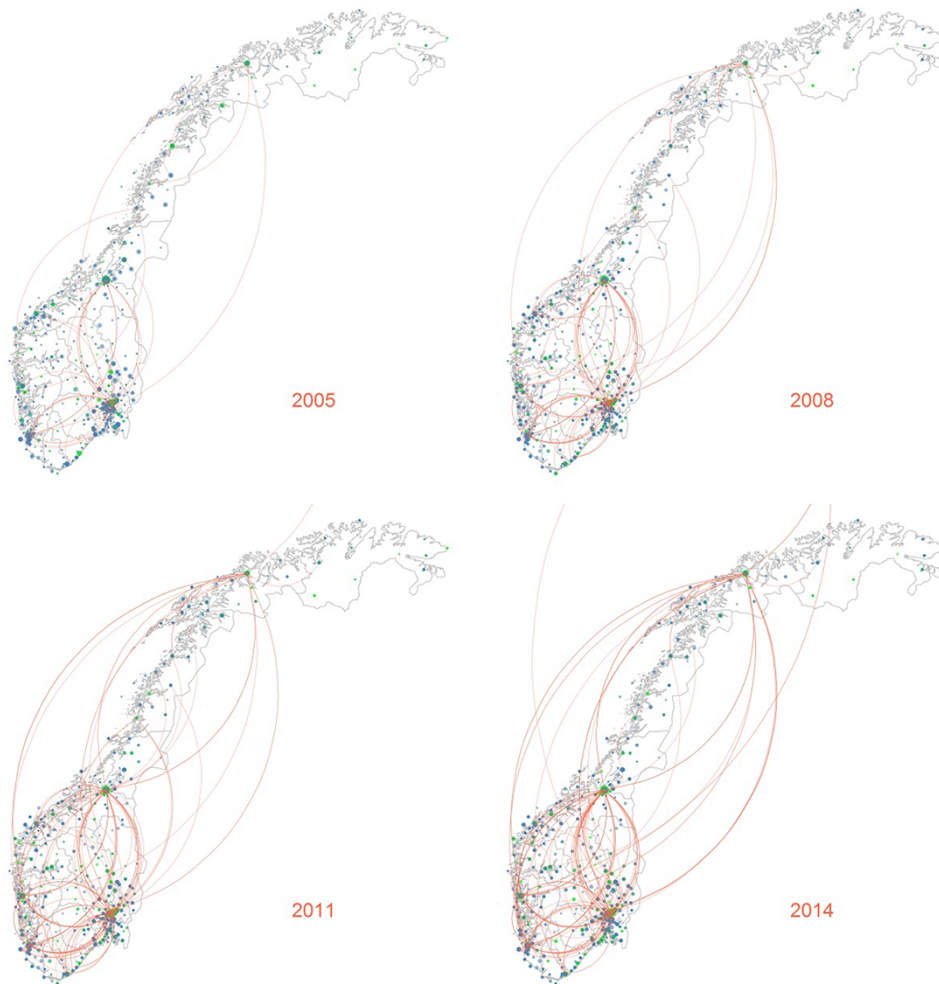
In recent years, the RCN has shifted its funding focus toward the long-term research and higher education plan developed by the Norwegian parliament (Regjeringen, 2018). The RCN serves under the Ministry of Education and Research, and half of the funding comes from this ministry and the Ministry of Trade and Industry.<sup>7</sup> The only specific industrial sectors mentioned in the long-term plan are the oil and gas, marine, and maritime sectors (Regjeringen, 2018). These are arguably Norway's three largest industrial sectors, and the Norwegian innovation system has been designed around them, as well as co-evolved with them (Reve and Sasson, 2012).<sup>8</sup> We also include the biotech sector in the analysis. This is a growing sector internationally, and a reference for sectors dominated by complex technologies. This sector is regarded as one of the highly innovative fields and has attracted a great deal of attention in recent years in Norwegian research funding (Kunnskapsdepartementet, 2011).

In Norway, the biotech sector and its related field of research are rather small compared to other countries. The industry mostly consists of small, innovative start-ups. The most substantial contribution of this sector to the Norwegian economy is joint projects in the aquaculture industry (The Research Council of Norway, 2016), where there is an increasing need for knowledge concerning fish health and environmental issues related to salmon farming.

The upstream oil and gas sector makes a very significant contribution to the Norwegian economy (accounting for 54% of total export in Norway between 2007 and 2021, see Figure 3). The sector includes not only the major companies that run the oil and gas fields but also the companies that supply them with equipment and services. Over the last 50 years, a substantial part of the Norwegian industry infrastructure has increasingly become directly or indirectly involved in oil and gas activities. The prosperity of the oil and gas sector depends on oil prices. Between 2005 and 2014, it was consistently in a good condition, with a slight shock during the financial crisis after 2008. However, in 2014, the price crashed and the industry faced a crisis, leading

<sup>7</sup> The purpose of the Norwegian Research Council is stated on its webpage: The Research Council provides advice on how and in which areas to target investments in Norwegian research efforts. We have been charged with strengthening the knowledge base and encouraging research that can help to solve the Grand Challenges. The Research Council of Norway (2018) works to add value to the research system by facilitating research that actors in the system could not successfully achieve working on their own.

<sup>8</sup> There is an extensive body of literature on the Norwegian innovation system—limited space prevents a full review in this paper. We have however applied these insights in the discussion part to explain the observed results within each sector. For a thorough review of the characteristics and historical development of the Norwegian national innovation system, see, e.g., Fagerberg *et al.* (2009).

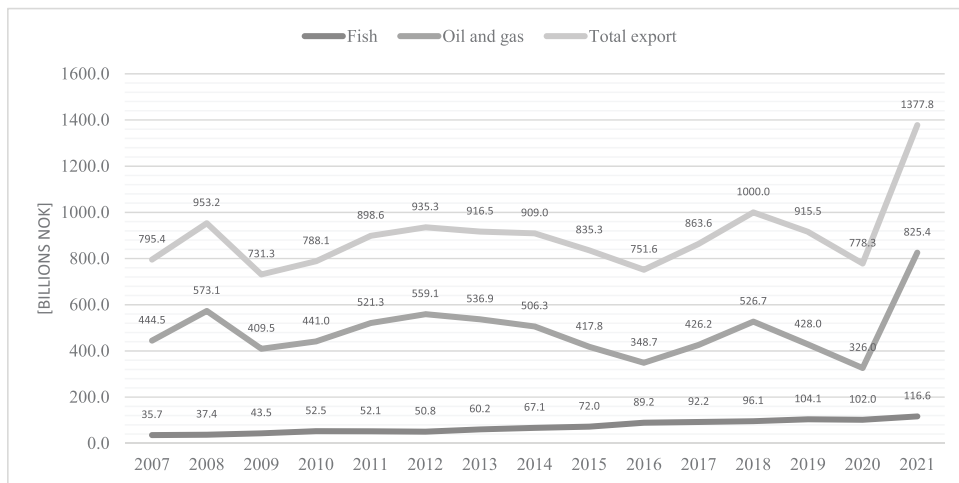


**Figure 2.** Position of actors and collaborative ties in R&D projects in Norway (only highly repeated collaborations are shown)

to the downscaling of all industries related to the oil and gas sector (Norsk Petroleum, 2018). Accordingly, the turmoil might have resulted in increased interest in long-term projects, due to companies' increased access to personnel and willingness to invest in alternative future markets. Similarly, applying for government funding for long-term research could be more attractive when internal funding is scarce.

We have chosen to use the name marine sector for the Norwegian industries that are involved in the production of seafood, including fisheries and aquaculture. This sector has experienced strong development in recent years, and its export value has tripled between 2007 and 2021 (see Figure 3). At the same time, this sector has encountered difficulties, particularly in salmon aquaculture. Issues with salmon lice, pollution, and escapes have put the sustainability of the industry into question and halted the growth of production volume (Hersoug, 2015). To maintain its reputation and competitiveness, the sector has tapped into broader knowledge sources to be able to deal with these issues, such as marine ecology, nutrition science, breeding, oil and gas engineering, artificial intelligence, medicine, and pharmaceuticals.

The modern maritime sector dates to the end of the 19th century. Maritime is the second-largest exporting sector in Norway (after oil and gas). The share of this sector in the Norwegian economy is significant and was reported to be stable between 2005 and 2015 (Menon, 2017). The maritime sector has increasingly supplied the Norwegian oil and gas industry with vessels



**Figure 3.** Export in Norway from 2007 to 2021 in billion NOK (based on Statistics Norway: <https://www.ssb.no/utenriksokonomi/utenrikshandel/statistikk/utenrikshandel-med-varer>, accessed at March 23, 2022)

and technical solutions. In fact, most of the growth in this sector is related to the oil and gas industry. As a result, it was adversely impacted when the oil price plummeted in 2014. Shortly after, one out of 10 jobs disappeared in the industry, which demonstrates how strongly this industrial sector is connected to the oil and gas sector (Menon, 2017: 3). Table 2 provides a brief overview of R&D projects in the biotech, oil and gas, marine, and maritime sectors across four time windows (descriptive statistics on the structural properties of each sector’s network provided in Appendix A).

### 3.3 Sectors’ complexity

The complexity literature offers multiple complexity measures ranging from economics to network science (Fleming and Sorenson, 2001; Hidalgo and Hausmann, 2009; Emmert-Streib and Dehmer, 2012; Tacchella *et al.*, 2012). We opt for the network density score (NDS) method developed by Emmert-Streib and Dehmer (2012) because contrary to other common measures of complexity such as the  $n/k$  model (Fleming and Sorenson, 2001) and the “method of reflections” (Hidalgo and Hausmann, 2009), the NDS is not biased by other attributes of technologies such as the geographic distribution of technologies and market conditions (Broekel 2019). This implies that networks with similar structural properties to a random network (Erdős and Rényi, 1960) have a higher degree of complexity compared to a more centralized network. Building on information theory, Broekel and Bednarz (2019: 4) argue that the NDS provides a reliable measure of technological complexity because “the more information is required to describe the topology of a technology’s combinatorial network, the more complex it is” (for a technical overview, see Appendix B).

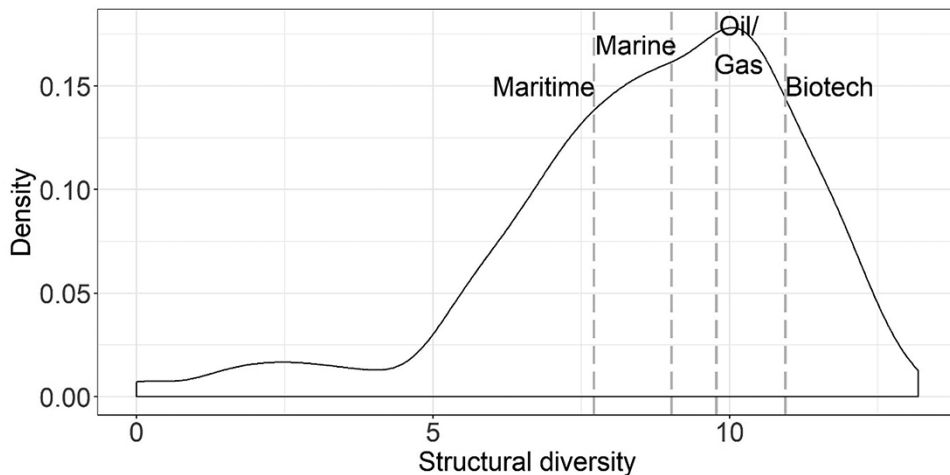
To estimate the complexity of sectors, we used the OECD REGPAT database (version: January 2020).<sup>9</sup> This database includes more than 3 million patents and provides information on the location of inventors, assignees, technological codes, and the filing and grant dates for each patent. We identified the main technologies for each sector using technological codes (for a list of technologies and their complexity, see Appendix B). We followed the method developed by Broekel (2019) to estimate the NDS for each technological code at the four-digit level. Since technological networks are dynamic phenomena, the NDS varies over time. Thus, for all technologies (four-digit level technology codes), we used the mean values of the NDS for three time windows (2005, 2010, and 2015) to minimize the effect of potential fluctuation in the data. Figure 4 shows

<sup>9</sup> For a detailed overview, see OECD (2008).

**Table 2.** Overview of the R&D database in Norway

Sector	Time window	Share of actors based on size <sup>a</sup>			Average number of actors per project	Number of projects	Actors in agglomeration (share)
		Public organizations (share)	Small	Medium			
Biotech	2005	0.249	0.47	0.157	0.373	2.087	0.645
	2008	0.231	0.516	0.186	0.299	1.498	0.614
	2011	0.223	0.523	0.185	0.293	1.216	0.596
	2014	0.216	0.531	0.19	0.279	1.148	0.573
Oil and gas	2005	0.184	0.324	0.213	0.464	1.882	0.792
	2008	0.18	0.395	0.215	0.39	1.303	0.765
	2011	0.157	0.429	0.216	0.355	1.286	0.717
	2014	0.146	0.46	0.203	0.338	1.279	0.708
Marine	2005	0.202	0.425	0.213	0.366	2.246	0.656
	2008	0.174	0.499	0.208	0.295	1.449	0.617
	2011	0.156	0.527	0.211	0.261	1.279	0.555
	2014	0.144	0.529	0.204	0.267	1.322	0.557
Maritime	2005	0.186	0.366	0.267	0.378	4.649	0.593
	2008	0.176	0.444	0.24	0.321	2.841	0.575
	2011	0.147	0.505	0.229	0.267	2.169	0.507
	2014	0.156	0.477	0.234	0.289	2.312	0.513

<sup>a</sup> Small actors: >50 employees; medium-sized actors: 50–250 employees; and large actors: >250 employees.



**Figure 4.** Complexity of sectors based on their dominant technologies. *Note:* Labels correspond to the median of the NDS for corresponding technologies

an approximation of the density of the NDS (kernel density estimation). The complexity of each sector is equal to the median of the NDS values for corresponding technologies. Among the four investigated sectors, biotech shows the highest (NDS: 10.94) and maritime the lowest (NDS: 7.71) degrees of complexity.

The estimated complexity score for sectors is in line with the intuitive knowledge of technological complexity of each category as well as two stylized facts associated with technological complexity. That is, (1) activities associated with complex technologies require greater R&D efforts, and (2) activities with complex technologies are more strongly concentrated in larger cities (Balland and Rigby, 2016; van der Wouden, 2020; Balland *et al.*, 2020b). Table 2 shows that the number of projects focusing on the biotech and oil and gas sectors (1303 and 1079 projects, respectively) is almost two times greater than the one of projects focusing on the maritime sector (610 projects). The high number of projects in the marine sector (2372 projects) should be viewed by its specificity and importance in Norway. Table 2 also reports the share of organizations located in large cities and their surrounding municipalities. R&D activities related to the biotech and oil and gas sectors have the highest share (share: 0.61 and 0.74, respectively), whereas maritime-related R&D activities less often occur in large cities (share: 0.55).

### 3.4 Variables

#### 3.4.1 Proximity variables

The concept of proximity can be very well expressed at the dyad level because proximity is the extent to which two economic agents are close at different levels or share a common property (Boschma, 2005). Arguably, scholars have mostly used the Euclidian distance between organizations to create a variable that captures the effect of geographical proximity (Micek, 2018). In a similar vein, the geographical proximity variable (GEO) corresponds to the log-transformed of the Euclidian distance between the centroids of NUTS regions (Nomenclature of Territorial Units for Statistics) at the 3-digit level (hereafter, NUTS3) in which organizations are located, multiplied by minus one.

Notwithstanding the common use of the Euclidian distance to define geographical proximity, it does not capture the nonlinear relation between geographical distance and knowledge sourcing (Micek, 2018). This is especially the case in Norway because 20% of the population live in the four largest cities, where most universities, firms, and research organizations are clustered. To tackle this problem, we distinguish between within and across agglomerations on knowledge sourcing.

In this light, SAME\_AGGLOM captures the effect of the colocation of organizations in an agglomeration on knowledge sourcing. This variable takes the value of one when two organizations are located in the same agglomeration,<sup>10</sup> and it takes the value of zero otherwise. Alternatively, DIF\_AGGLOM captures non-local interactions between four large agglomerations in Norway. It takes the value of one for collaborating actors located in different agglomerations and the value of zero otherwise. ONE\_AGGLOM is a dummy variable that distinguishes between collaborations with one collaborator located in an agglomeration and one collaborator not located in one of the four agglomerations (taking the value of one). By doing so, the locations of both collaborators in non-agglomeration areas serve as a baseline variable, compared to which the effect of other geographical proximity dummy variables can be interpreted.

Cognitive proximity is the extent to which two actors are capable of understanding and interpreting exchanged information (Cohen and Levinthal, 1990). The Norwegian R&D database includes data on the domains of activities for each project. Building on this, one can disaggregate these domains of activities and assign them to each project participant. This provides cognitive portfolios for all participants. The cognitive portfolio is a function of the intensity and diversity of projects, in which a given organization has been involved. The variable capturing the effect of cognitive proximity (COG) is based on the Jaccard index for similarity of each two organizations' cognitive portfolios. This index is defined as follows:

$$\text{COG} = \frac{|\text{ORG}_i \cap \text{ORG}_j|}{|\text{ORG}_i \cup \text{ORG}_j|} \quad (1)$$

where  $\text{ORG}_i$  and  $\text{ORG}_j$  refer to two vectors including domain tags of projects, in which organizations  $i$  and  $j$  have been involved. COG ranges from zero to one. COG takes the value of zero if two organizations do not share a common domain tag and takes the value of one if two organizations have been involved in projects with the same tags. This definition of cognitive proximity is of critical importance because this measure does not rely on binary similarity measures (e.g., NACE codes) and takes inter-sectoral dependencies into consideration.

The institutional proximity variable (INST) is based on the commonalities and differences in the routines and procedures of organizations. Broekel and Boschma (2012) argue that public and private organizations differ regarding their objectives and strategies. INST is a dummy variable, it takes the value of one if both organizations in an organization pair are either privately (e.g., private companies) or publicly (e.g., universities) funded, and it takes the value of zero otherwise.

This is important to note that this study does not include a variable that captures the effect of social proximity. Given the size of the secondary dataset (1765 organizations) used for this study, we are not able to identify whether companies share a common past or whether there are friendship or kinship relations between employees that work in different companies. Also, the dataset only includes information about collaborative ties between organizations and not about the ones within organizations. Therefore, we refrained from creating a variable that captures the effect of organizational proximity.

### 3.4.2 Node-level variables

These variables were designed to control for individual effects that might bias the results of the econometric models. To define a proxy for the size of organizations, we created SIZE based on a widely used method. SIZE is a categorical variable that takes four values:

- Category 1 (micro): <6 employees,
- Category 2 (small): 6–50 employees,
- Category 3 (medium): 51–250 employees,
- Category 4 (large): >250 employees.

Besides the size of organizations, two dummy variables control for the type and structure of organizations. PUBLIC takes the value of one if an organization is fully publicly funded and takes

<sup>10</sup> Agglomerations are the four largest cities in Norway (Oslo, Bergen, Trondheim, and Stavanger) in Norway with their surrounding municipalities.



**Table 3.** Descriptive statistics of the variables

Statistic	N <sup>a</sup>	Mean	St. Dev.	Min.	Pctl(25)	Pctl(75)	Max.
GEO	1,556,730	0.00	1.00	-1.41	-0.54	0.27	2.89
COG	1,556,730	0.00	1.00	-0.86	-0.86	0.83	2.03
INST	1,556,730	0.66	0.47	0	0	1	1
SAME_AGGLOM	1,556,730	0.10	0.30	0	0	0	1
DIF_AGGLOM	1,556,730	0.22	0.42	0	0	0	1
ONE_AGGLOM	1,556,730	0.49	0.50	0	0	1	1
SIZE_origin	1,556,730	2.63	1.03	1	2	4	4
SIZE_destination	1,556,730	2.32	1.01	1	2	3	4
PUBLIC_origin	1,556,730	0.17	0.38	0	0	0	1
PUBLIC_destination	1,556,730	0.25	0.44	0	0	1	1
HYBRID_origin	1,556,730	0.03	0.17	0	0	0	1
HYBRID_destination	1,556,730	0.03	0.17	0	0	0	1
AGGLOM_origin	1,556,730	0.44	0.50	0	0	1	1
AGGLOM_destination	1,556,730	0.43	0.50	0	0	1	1

<sup>a</sup>The number of observations corresponds to the number of organization pairs (i.e.,  $n \times (n - 1) / 2$ , where  $n$  denotes the number of organizations.  $n$ : 1765).

the value of zero otherwise. Given the specificity of the Norwegian context, it is plausible that the government owns companies. To control for this, HYBRID is the second dummy variable that takes the value of one if a company is partly or fully owned by the government, and it takes the value of zero if the share of the government ownership is zero.

The importance of agglomeration in terms of skilled labor pooling (Marshall, 1920), competitiveness (Porter, 1998), and labor division (Jacobs, 1970) has been widely discussed. These factors are of great importance when it comes to knowledge sourcing (Carlino and Kerr, 2015). AGGLOM is a dummy variable that takes the value of one when a given organization is located in an agglomeration and takes the value of zero otherwise. Table 3 shows descriptive statistics for the variables (see Appendix C for the correlation matrix).

### 3.5 Model construction

Researchers have multiple analytic tools for analyzing relational data by means of inferential statistics (for a review, see Broekel *et al.*, 2014). Since the hypotheses focus on the effect of proximity dimensions, spatial interaction models provide potent analytic tools, whereby one can specify and control for the effect of the key factors at the node (control variables) and dyad levels (proximity dimensions). The gravity model was originally derived from Newton's law of universal gravitation (for a detailed review of the model, see Head and Mayer, 2014). In the 1960s, this model became popular among those who aimed to analyze trade between countries (e.g., Hasson and Tinbergen, 1966). The model implies that geographically close and large economies trade more often compared to distant and smaller ones. The basic model is defined as follows:

$$X_{ij} = K \frac{Y_i^{\beta 1} Y_j^{\beta 2}}{d_{ij}^{\beta 3}} \quad (2)$$

where  $X_{ij}$  represents the presence or intensity of the interaction between two objects  $i$  and  $j$ ;  $Y$  is the size of the objects;  $d_{ij}$  denotes the distance between the objects; and  $K$  is a constant.  $\beta 1$  and  $\beta 2$  are the estimated parameters that correspond to the intensity of generated and attracted interactions, respectively.  $\beta 3$  corresponds to the impeding effect of distance between objects  $i$  and  $j$  (Broekel *et al.*, 2014). Since the gravity model has a multiplicative form, one can take the logarithm of both sides and transform Equation (2) into a linear form.  $Z$  and  $u_{ij}$  denote any hidden factors at the node and dyad levels:

$$\ln X_{ij} = \ln K + \beta 1 \ln Y_i + \beta 2 \ln Y_j - \beta 3 \ln d_{ij} + Z + u_{ij} \quad (3)$$

Using the gravity model provides the possibility to include multiple variables at the node (e.g., size) and the dyad levels (e.g., geographical proximity). Also, compared to other dominant methods (e.g., the exponential random graph model), the gravity model ensures better leverage and reliability when the observation number exceeds a certain threshold.<sup>11</sup>

To estimate the gravity model, we apply a set of pooled logistic regression models. The dependent variable is a dummy variable that takes the value of one if there is a collaborative tie between two given organizations and takes the value of zero otherwise.<sup>12</sup> Organization pairs are the units of analysis in this study.<sup>13</sup> The model is defined as follows:

$$\ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \beta 1 \text{Geographic}_{ij} + \beta 2 \text{COG}_{ij} + \beta 3 \text{INST}_{ij} + \beta 4 \text{Node}_i + \beta 5 \text{Node}_j + \varepsilon + \omega_i + \omega_j \quad (4)$$

where  $P_{ij}$  denotes the probability of the presence of a collaborative tie between organizations  $i$  and  $j$ .  $\text{Geographic}_{ij}$  corresponds to geographical proximity variables (GEO, SAME\_AGGLOM, DIF\_AGGLOM, and ONE\_AGGLOM) between organizations  $i$  and  $j$ . Similarly,  $\text{COG}_{ij}$  and  $\text{INST}_{ij}$  denote variables capturing the effect of cognitive and institutional proximities. Finally, Node represents a set of control variables at the node level (i.e., SIZE, PUBLIC, HYBRID, and AGGLOM) associated with organizations  $i$  and  $j$  (origin and destination).  $\omega_i$  and  $\omega_j$  represent two categorical fixed effects (FEs) for organization types. Organization types consist of the following categories: (1) association and industry organizations, (2) banks, (2) private firms and companies, (4) county-level administrations, (5) government agencies, (6) independent research institutes, (7) municipalities, (8) nongovernmental organizations (NGOs), (9) public hospitals and health organizations, (10) public–private cluster organizations, (11) schools and museums, (12) university and university colleges (“høyskole”), and (13) others.  $\varepsilon$  denotes the sum of error terms at the node and dyad levels introduced in Equation (3) (i.e.,  $\varepsilon = Z + u_{ij}$ ).

By adding an interaction term, this study also investigates the multiplicative nature of two variables of concern to capture the complementarity and substitutability of the effects of cognitive proximity and colocation in large cities. Besides the interaction term, the model setting is similar to Equation (4).

#### 4. Results and discussion

This section focuses on the results of the gravity model operationalized by several logistic regressions. We conducted regression models in a stepwise manner for each sector. In Model 1, we introduced organization-type FEs and control variables (i.e., SIZE, PUBLIC, HYBRID, and AGGLOM). In Model 2, we added variables that capture the effects of organizations' location in the same and different agglomerations (SAME\_AGGLOM, DIF\_AGGLOM, and ONE\_AGGLOM). Finally, in the Full Model, we also included all the three proximity variables (i.e., GEO, COG, and INST). Appendix D provides four regression tables (for each investigated sector: biotech, oil and gas, marine, and maritime) that report the coefficients of Models 1, 2, and Full Model. Figure 5 illustrates that Full Models for all sectors provide the best goodness of fit, while the sign and the significance of reported coefficients do not change (see Appendix D). Thus, we interpret and discuss the results of Full Models.

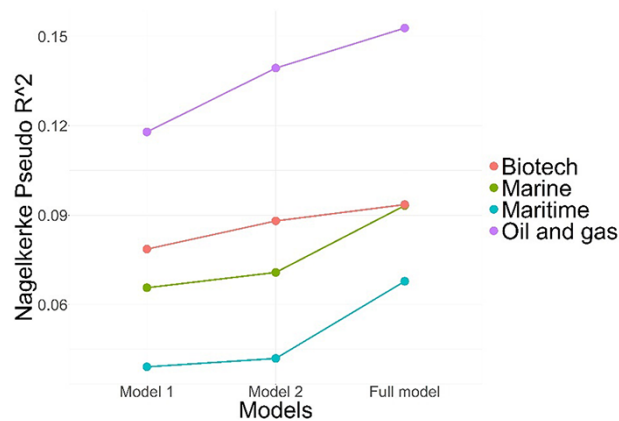
Table 4 provides the results. Each model demonstrates the association of proximity dimensions and control variables with collaborative tie formation in each sector. The sign and significance of control variables mostly support our expectation and resonate with other empirical studies.

First, the size of organizations is positively correlated with the likelihood of collaborative tie formation. Second, organizations that are in four large agglomerations are more likely to create R&D ties in all cases except in the oil and gas sector. This might have to do with the geographic

11 Normally, there are persisting convergence problems in Exponential Random Graph Model (ERGMs) and Stochastic Actor Oriented Models (SAOMs) if the number of nodes (e.g., organizations) exceeds 1000. This study includes 1765 organizations.

12 All continuous variables are log-transformed.

13 The number of observations corresponds to  $(n \times (n - 1))/2$ , where  $n$  is the number of organizations (1765). Out of 1,556,730, 77 organization pairs were removed due to missing data.



**Figure 5.** Goodness of fits across models

distribution of organizations active in the oil and gas sector. Third, we observe varying results regarding the effect of the type of organizations on the likelihood of being involved in R&D projects. While public organizations (non-companies) tend to be relatively more active in tie formation in the oil and gas and maritime sectors, this variable is found to be negatively correlated in the marine and biotech sectors. This finding arguably reflects the different nature of these four sectors. The oil and gas and maritime sectors include to a larger extent publicly funded organizations and university departments. Interestingly, the results suggest that collaborative ties are more likely to be established between public and private organizations in the marine sector. Fourth, the likelihood of creating collaborative ties increases if the government partly or fully owns the organizations. This effect is larger for the oil and gas and marine sectors. Lastly, institutional proximity is found to be positively correlated with the likelihood of tie formation in the marine, biotech, and oil and gas sectors.

Having briefly reported the results regarding the control variables, we mainly discuss the results in relation to the hypotheses in this section, and how the reported coefficients and their signs, significance, and effect sizes support or reject the hypotheses.

Geographical proximity is one of the main variables of concern in this paper. Interestingly, we do not observe that geographically close organizations are more likely to involve in joint R&D projects. In the case of the maritime sectors, geographical proximity is even negatively correlated with the likelihood of collaborative ties being created. Conversely, the effect of the colocation of two organizations in an agglomeration<sup>14</sup> is found to be positive and statistically significant across all sectors. The result partly supports Hypothesis 1. This finding resonates with [Sonn and Storper's \(2008: 1021\)](#) argument: “[...] time and costs are not linearly related to distance. Rather, they resemble a step function in which marginal costs rise very steeply once an overnight step is required. [...] distance should therefore be defined as a discontinuous variable with distinctive thresholds values.” This implies that firms only enjoy the geographical proximity within an agglomeration, whereas this is not relevant for the ones located outside the boundaries of one agglomeration. In addition, it is plausible that the positive effect of SAME\_AGGLOM is partly driven by other agglomeration spillover effects. This could also imply that firms and institutes located in peripheral areas do not have the required absorptive capacity ([Cohen and Levinthal, 1990](#)) to establish collaborative ties with those in large cities. This is of crucial importance for innovation policy, since the landscape of innovation has become increasingly spiky, bringing about spatial and consequently social inequalities ([Florida, 2005](#)). This claim is supported by the result reported for the other two dummy variables related to agglomeration effects. That is, the

<sup>14</sup> As explained before, the baseline variable for the effect of the colocation of two organizations in an agglomeration is the colocation of both organizations in non-agglomeration regions, and thus, the effect of this variable should be interpreted compared to the baseline variable.

Table 4. Results of the regression models

	Marine	Maritime	Biotech	Oil and gas
COG	0.480*** (0.009)	0.553*** (0.013)	0.268*** (0.015)	0.405*** (0.014)
SAME_AGGLOM	0.682*** (0.034)	0.154*** (0.058)	0.246*** (0.065)	1.653*** (0.048)
COG × SAME_AGGLOM	0.014 (0.021)	-0.073*** (0.035)	0.151*** (0.033)	0.216*** (0.031)
DIF_AGGLOM	0.067* (0.032)	-0.418*** (0.055)	-0.694*** (0.066)	1.198*** (0.046)
ONE_AGGLOM	-0.137*** (0.021)	-0.256*** (0.034)	-0.429*** (0.039)	-0.048 (0.032)
GEO	-0.009 (0.010)	-0.041*** (0.015)	-0.007 (0.015)	0.004 (0.016)
INST	0.200 (0.023)	0.050 (0.036)	0.110 (0.032)	0.152 (0.037)
SIZE_origin	0.429 (0.011)	0.371 (0.016)	0.303 (0.019)	0.679 (0.020)
SIZE_destination	0.474 (0.010)	0.380*** (0.016)	0.291*** (0.017)	0.757 (0.017)
PUBLIC_origin	0.207*** (0.056)	0.725*** (0.100)	-0.218*** (0.100)	0.747*** (0.085)
PUBLIC_destination	-0.238*** (0.068)	0.179 (0.116)	-0.712*** (0.111)	0.282 (0.095)
HYBRID_origin	0.620 (0.041)	0.270*** (0.072)	0.569*** (0.078)	1.082 (0.052)
HYBRID_destination	0.388*** (0.041)	-0.002 (0.075)	0.098 (0.078)	0.584*** (0.062)
AGGLOM_origin	0.058 (0.025)	0.207*** (0.039)	0.500*** (0.044)	-0.236*** (0.038)
AGGLOM_destination	0.072 (0.025)	0.209*** (0.039)	0.531*** (0.044)	-0.274*** (0.038)
Constant	-7.584*** (0.126)	-8.953*** (0.208)	-6.096*** (0.187)	-13.546*** (0.278)
Organization-type FE	Yes	Yes	Yes	Yes
Observations <sup>a</sup>	1,556,730	1,556,730	1,556,730	1,556,730
Nagelkerke's R <sup>2</sup>	0.093	0.068	0.094	0.153
Log-likelihood	-58,329.030	-29,393.960	-26,275.240	-24,292.250

\*  $P < 0.1$ .\*\*  $P < 0.05$ .\*\*\*  $P < 0.01$ .<sup>a</sup> The number of observations corresponds to the number of organization pairs (i.e.,  $n \times (n - 1)/2$ ), where  $n$  denotes the number of organizations.  $n$ : 1765).

likelihood of involving in a joint R&D project drops when one organization is in an agglomeration and the other one not (see Table 4). This is to a lesser extent the case for organizations that are in different agglomerations.

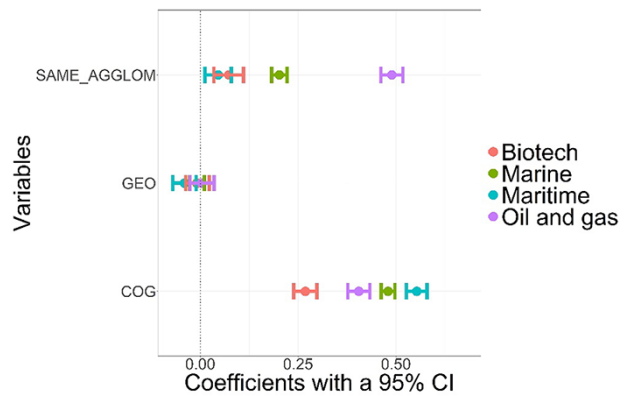
While our findings support Hypothesis 2, the results could be partly influenced by the geographical particularities of Norway where large cities are relatively far from one another and there is a high degree of concentration of headquarters and public research institutes in them. This finding may also point toward a policy effect, and the attributes of a given national innovation system as an alternative factor, that increases the relevance of colocation in agglomerations compared to the one of geographical proximity. Thus, it seems necessary that future studies test this hypothesis in other contexts.

Cognitive proximity is of great interest when it comes to sectoral dynamics. The results show that the cognitive proximity is positively related to the likelihood of involving in joint R&D projects. The biotech sector is perhaps a prime example of a knowledge-intensive sector (Ter Wal, 2014). While biotech is already an established sector and getting involved in R&D projects might be the most common strategy among organizations in this field (Owen-Smith and Powell, 2004), the underlying knowledge base is complex and involvement in joint projects requires a certain degree of field-related knowledge and absorptive capacity. This finding is consistent with the conceptual framework developed by Ter Wal and Boschma (2011), who argue that cognitive proximity plays a crucial role in establishing collaborative ties in growing industries.

The oil and gas sector represents the largest industrial sector in Norway (Prestmo *et al.*, 2015), with a diverse range of actors collaborating to develop technologies for extracting offshore oil and gas. This sector functions mainly through collaborative engineering projects involving heavy equipment, massive machine tools, and the related infrastructure. Since this sector has the characteristics of a mature industry with a high degree of knowledge codification, one might expect that cognitive proximity does not play a crucial role in interorganizational tie formation (Cowan *et al.*, 2004). However, this sector has experienced considerable turbulence over the years that the database covers. It is plausible that the structure and maturity of the Norwegian continental shelf has called for more advanced petroleum extraction methods (Engen *et al.*, 2018). This has increased the need for new technologies in this sector and led to an increasing number of interactions among actors with similar cognitive portfolios.

Although the marine sector is one of the oldest in Norway, the importance of cognitive proximity might be the result of the change in the industry toward more advanced vessels and a large increase in salmon farming. There has been an exponential increase in demand for Norwegian fishing products and substantial changes in the composition of actors in the R&D network. The growth of salmon farming (now a substantial part of this industry in Norway) has resulted in a wide range of industry-specific challenges. The boom in farmed fish has led to an increased focus on issues related to the impact on the wild salmon population, local environmental issues in coastal waters, and alarming problems associated with salmon lice and related diseases. Because salmon farming is one of Norway's core industries, government actors have targeted these challenges in recent years. The upturn in the industry has, coupled with stricter governmental regulations, forced actors to adapt their strategies to confront new challenges (Sarpebakken and Ubisch, 2017). This has also motivated a larger number of firms and organizations (partly subsidized by the RCN) to collaborate with the biotech sector in tackling the challenging issues (Hersoug, 2015). Examples of such collaborations are marine biotechnology and health-related research communities that focus on solving the disease and environmental issues. Given the ever-increasing complexity of cutting-edge technologies, these projects require the teamwork of experts with small cognitive distances, that is, experts from the same or related fields (Nooteboom, 2000; Hidalgo, 2015; Balland and Rigby, 2016).

The maritime sector shows the largest effect size regarding the cognitive variable. That is, sharing similar cognitive portfolios increases the odds of tie formation the most, compared to the other three sectors. This might be driven by the fact that maritime-related companies mostly collaborate with more similar actors to develop technologies for their core market. Many companies have traditionally provided services to the oil and gas industry. In recent years, these companies have been involved in developing solutions for the aquaculture industry and renewable energy



**Figure 6.** Standardized coefficients with corresponding 95% confidence intervals reported for cognitive proximity, geographical proximity, and colocation in an agglomeration

projects. This involves applying knowledge about sea structures to larger and more complex fish farms further out to sea and offshore wind farms (Mäkitie *et al.*, 2018).

Notably, the standardized coefficients reported for cognitive proximity are higher than the ones for geographical proximity and colocation in an agglomeration (Figure 6). The only exception is observed in relation to the oil and gas sector. The relatively higher effect size of the location in the same agglomeration in the oil and gas sector might be driven by the headquarters effect because a large share of Norwegian oil and gas headquarters are clustered in Oslo and Stavanger (0.44) compared to the ones of other sectors (biotech: 0.28, marine: 0.25, and maritime: 0.29). These results taken together partly support Hypothesis 3.

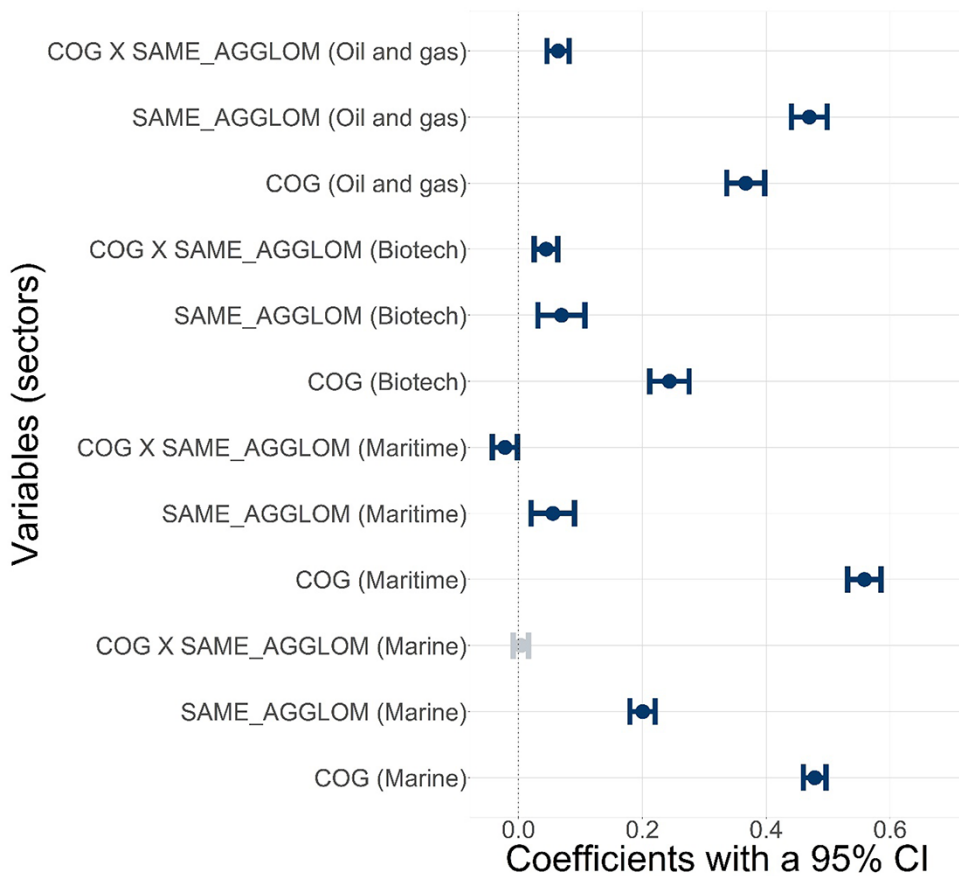
Table 4 also provides results in relation to the joint effects of cognitive proximity and colocation in the same agglomeration. We reported two models for each sector with and without the interaction term to ensure that including the new variable does not impact the sign and the significance of the variables of concern.

Figure 7 provides the standardized regression coefficients. Notably, while the effects of cognitive proximity and colocation in agglomerations are positively related to collaborative tie formation, we observe that the sign of the joint effects of these two variables varies across sectors. That is, the joint effects are negative and statistically significant for the maritime sector, whereas this is positive for the other three sectors and statistically significant only for the oil and gas and biotech sectors. This provides evidence to support the hypotheses regarding the complementarity and substitutability between the effects of cognitive proximity and colocation in large cities (Hypotheses 4 and 5). The maritime sector is identified as a sector with relatively lower complex technologies. The negative joint effects between cognitive proximity and colocation in large cities might point toward the fact that these two factors substitute each other's effects. The complexity of the technologies associated with the marine sector (9.02) is slightly higher than the average complexity of all technologies (8.99). In this sector, it seems that the two variables behave independently from one another. Yet, the effects of two variables of concern complement one another perhaps because the biotech and oil and gas sectors are associated with technologies with a higher degree of complexity (10.94 and 9.78, respectively). Our finding does not concur with the argument developed by Balland *et al.* (2020b) who claim that complex activities concentrate in large cities perhaps because the geographical proximity could compensate for the lack of the cognitive proximity. We observe, however, that organizations active in sectors dominated by complex technologies ideally require both cognitive proximity and colocation in large cities (i.e., complementarity effects).

#### 4.1 Robustness checks

To ensure the robustness of our results, we conducted additional models with alternative variables and model specifications. First, since the average duration of R&D projects is 3.09 years, we ran





**Figure 7.** Results of regression models with interaction terms

models for four time periods (3-year time windows) to ensure that the sign and significance of the variables of interest do not change over time. The reported coefficients of 32 regression models (4 time windows  $\times$  4 sectors  $\times$  2 with and without interaction terms) are in line with the ones of the pooled models.

Second, to ensure that we exclude false-positive results in the models with an interaction term, we estimated a set of models in which the variable capturing the effect of cognitive proximity is a dummy variable. The dummy variable takes the value of one if the value of the original cognitive proximity variable is greater than the 90th percentile, otherwise it takes the value of zero. The reported outcome is also robust against this alternative specification (the results are available upon request from the authors).

Third, we estimated the gravity model by following the work of [Hoekman \*et al.\* \(2009\)](#) who included two variables at the importer and exporter levels. For instance, we included two variables that capture the effects of the size of organizations on the collaborative tie formation (i.e., `SIZE_origin` and `SIZE_destination`). Alternatively, the classical works ([Isard, 1956](#); [Tinbergen, 1962](#)) modeling socioeconomic phenomena using the gravity models that include one variable that represents the multiplicative nature of the origin- and destination-level variables (e.g., `SIZE` which corresponds to `SIZE_origin`  $\times$  `SIZE_destination`). Similar results hold when we ran alternative models with an alternative size variable ([Appendix E](#)). It is important to note that in this robustness check, we excluded the organization-level variables whose multiplication was highly correlated with dyadic proximity variables.

Fourth, to enhance the interpretability of results, the number of observations is constant across models that estimate the effects of different variables on the likelihood of collaborative tie formation in the four sectors. This implies that as the number of organizations increases, the number

of potential relations scales more quickly compared to the one of observed ties. This causes the rare event problem, and consequently the fitted model might suffer from small-sample bias (King and Zeng, 2001). To remedy this situation, we applied the Firth method (also known as penalized likelihood) to estimate new models to reduce the potential small-sample bias (Heinze and Schemper, 2002; Puhf *et al.*, 2017). The main results remain consistent with the ones of original models (Appendix F).

Finally, we aimed at estimating mixed-effects regression models. While researchers often estimate gravity models with a fixed-effect approach, this does not include estimations with random exporter and importer intercepts (i.e., organization pairs in this work) (Prehn *et al.*, 2015). We followed the suggested approach by Prehn *et al.* (2015) and the R package lme4 (Bolker, 2014) to fit a set of mixed-effects models with random exporter and importer intercepts. The glmer algorithm could not estimate the mixed-effects models given the high number FEs and large dataset.

## 5. Conclusion

The aim of the present research was twofold. First, to address a conceptual issue in the sectoral system of innovation literature (Malerba, 2009), whereby the multi-layered interdependencies of sectoral actors in knowledge sourcing cannot be empirically observed. Following Boschma (2005), we postulate that knowledge sourcing is embedded in proximity dimensions within and across sectors. Second, to conduct an empirical work to gain a better understanding of how proximity dimensions are related to knowledge sourcing in four sectors in Norway. While proximity literature has received a great deal of attention, most empirical studies have been limited to geographical areas (e.g., Molina-Morales *et al.*, 2015; Lazzeretti and Capone, 2016; Belso-Martínez *et al.*, 2017). Our study provides empirical evidence on how sectors, proximities, and their joint effects interact. While proximity dimensions are similarly related to knowledge sourcing across sectors, the joint effects of cognitive proximity and colocation in large cities vary in sectors to a considerable degree. We conjecture that this variance might be driven by the complexity of underlying technologies in each sector, whereas it is difficult to account for such findings merely from the rationale behind Pavitt's (1984) sectoral patterns.

The notion of complementarity and substitutability of proximities is of crucial importance for innovation policy because one dimension can be replaced or reinforced by the other to increase the likelihood of creating collaborative ties. Indeed, Boschma and Balland (2020) show that inter-regional linkages can be used as a source of required capabilities. Our study points toward the sectoral specificity of knowledge sourcing. Ideally, it seems that organizations active in sectors dominated by relatively complex technologies require both colocation in large cities and cognitive proximity for involving in joint projects. This begs the pressing question of what adjustments are needed to ensure the efficiency of innovation policy measures that aim at promoting collaborations among cognitively distant actors. This is especially the case in the context of mission-oriented policies targeted by multiple European countries (Mazzucato, 2018). The aim of such policies is to motivate and coordinate interdisciplinary collaborations between a wide range of actors to find solutions for pressing societal challenges (Janssen and Abbasiharofteh, 2022; Wanzenböck *et al.*, 2020; Wanzenböck and Frenken, 2020). Our findings thus raise this question: how can mission-oriented policy encourage cognitive boundary-spanning knowledge sourcing behaviors in sectors dominated by complex technologies?

Of course, the shortcomings of this study need to be addressed by future research. While our study provides evidence that colocation in large cities facilitates the formation of collaborative ties, we acknowledge that the results have limited interpretability because colocation in cities is associated with various types of agglomeration spillovers. This motivates future studies that aim at disentangling the effects of existing forces that drive the formation of collaborative ties. In addition, we investigated a subset of the R&D dataset supported by the RCN that includes projects in biotech, oil and gas, marine, and maritime sectors. To ensure that our results hold for all other sectors, new studies can retrieve and disambiguate the complete dataset of the Norwegian R&D projects. Mapping the complete R&D network also enables researchers to ask theoretically informed questions about the relevance of structural properties of R&D networks in collaborative tie formations and to compare the results with the ones of studies that used

patent, scientific publication, and web-mining data (Breschi and Catalini, 2010; DeStefano and Zaccarin, 2013; Lata *et al.*, 2015; Abbasiharofteh *et al.*, 2020, 2021; Tóth *et al.*, 2021). The present research modeled interorganizational relations at the dyad level and disregarded tie formation mechanisms forced by the structural properties of a collaboration network (Broekel and Bednarz, 2019; e.g., Abbasiharofteh and Broekel, 2020). Applying more advanced inferential network models (e.g., Exponential Random Graph Models (ERGMs)) to a case study of this scale is computationally intensive and requires the further advancement of models. The advent of new statistical network models paves the way for understanding the impact of the structural properties of a collaboration network on the decision of sector actors in forming future collaborative ties. We use project tags manually added by the Research Council and there is of course suspect to human error, but categorizing projects or companies is in general a human process and we have no reasons to suspect that such tags should be less precise than other categorizations.

Also, Norway has a unique economic structure and geography, and therefore the findings might not be generalizable for all countries. However, the studied sectors are present in many places in the world, and several European countries have similar political structures and innovation systems compared to Norway. For instance, the foundations of the R&D system (Gulbrandsen and Nerdrum, 2009: 61) and the share of state-owned companies<sup>15</sup> are both similar to many other Western European countries (but rather different from the UK and the USA). Hence, we argue that the mechanisms found for forming collaborative ties in Norway might be generalizable to countries that are similar along such dimensions. In addition, the findings are relevant for the understanding of the four sectors analyzed in this study. That said, to ensure the external validity of our findings, more research in various contexts is needed. Future studies can thus use our work as a point of departure to study how the relevance of geographical and cognitive proximity dimensions varies with the complexity of sectors, as well as how various policy schemes co-evolve with such effects in national and sectoral innovation systems.

Lastly, scholars argue that the impact of proximity might change over time, and thus, the impact of proximities on knowledge sourcing should be studied in the long run (Balland *et al.*, 2020a). This evolutionary approach also contributes to a better understanding of why the effects of proximities complement or substitute one another. Yet, a precise understanding of the evolution of proximity dimensions is lacking. Given the structure of the data, we covered 12 years of R&D activities in four sectors. The ever-increasing abundance of relational data covering a longer period of time (e.g., see Petralia *et al.*, 2016) will enable future research to include longer time periods and thus capture the transition phase of industries.

## Funding

The Research Council of Norway, grant: 237677. Milad Abbasiharofteh acknowledges the support of the BMBF junior research group TRAFÖBIT (031B0020).

## References

- Abbasiharofteh, M. (2020), 'Endogenous effects and cluster transition: a conceptual framework for cluster policy,' *European Planning Studies*, 30, 1–24.
- Abbasiharofteh, M. and T. Broekel (2020), 'Still in the shadow of the wall? The case of the Berlin biotechnology cluster,' *Environment and Planning A: Economy and Space*, 46, 3.
- Abbasiharofteh, M., J. Kinne and M. Krüger (2021), 'The strength of weak and strong ties in bridging geographic and cognitive distances,' *ZEW Discussion Paper No. 21-049*, Mannheim.
- Abbasiharofteh, M., D. F. Kogler and B. Lengyel (2020), 'Atypical combination of technologies in regional co-inventor networks,' *Papers in Evolutionary Economic Geography (PEEG)* 20.55. <http://econ.geo.uu.nl/peeg/peeg2055.pdf> (accessed January 15, 2021).
- Archibugi, D. (2006), 'Pavitt's taxonomy sixteen years on: a review article,' *Economics of Innovation and New Technology*, 10(5), 415–425.
- Asheim, B. T. and L. Coenen (2005), 'Knowledge bases and regional innovation systems: comparing Nordic clusters,' *Research Policy*, 34(8), 1173–1190.

15 For a review, see <https://www.oecd.org/corporate/oecd-dataset-size-composition-soe-sectors.htm>

- Asheim, B. T., H. L. Smith and C. Oughton (2011), 'Regional innovation systems: theory, empirics and policy,' *Regional Studies*, 45(7), 875–891.
- Audretsch, D. and M. P. Feldman (1996), 'R&D spillovers and the geography of innovation and production,' *American Economic Review*, 86, 630–640.
- Balland, P.-A., J. A. Belso-Martínez and A. Morrison (2015a), 'The dynamics of technical and business knowledge networks in industrial clusters: embeddedness, status, or proximity?,' *Economic Geography*, 92(1), 35–60.
- Balland, P.-A., R. Boschma and K. Frenken (2015b), 'Proximity and innovation: from statics to dynamics,' *Regional Studies*, 49(6), 907–920.
- Balland, P.-A., R. Boschma and K. Frenken (2020a), *Proximity, Innovation and Networks: A Concise Review and Some Next Steps*. Department of Human Geography and Spatial Planning, Group Economic Geography, Utrecht University: Utrecht.
- Balland, P.-A., M. de Vaan and R. Boschma (2013), 'The dynamics of interfirm networks along the industry life cycle: the case of the global video game industry, 1987-2007,' *Journal of Economic Geography*, 13(5), 741–765.
- Balland, P.-A., C. Jara-Figueroa, S. G. Petralia, M. P. A. Steijn, D. L. Rigby and C. A. Hidalgo (2020b), 'Complex economic activities concentrate in large cities,' *Nature Human Behaviour*, 4(3), 248–254.
- Balland, P.-A. and D. Rigby (2016), 'The geography of complex knowledge,' *Economic Geography*, 93(1), 1–23.
- Bathelt, H. and J. Glückler (2018), 'Relational research design in economic geography,' in G. L. Clark, M. P. Feldman, M. S. Gertler and D. Wójcik (eds), *The New Oxford Handbook of Economic Geography*. Oxford University Press: Oxford, pp. 179–195.
- Belso-Martínez, J., M. Expósito-Langa, F. Mas-Verdú and F. Molina-Morales (2017), 'Dynamics of brokerage positions in clusters: evidence from the Spanish foodstuffs industry,' *Sustainability*, 9(2), 290.
- Bernela, B., M. Ferru and A. Rallet (2019), 'The impact of digital technologies on perceptions of proximity,' *hal-02051751f*.
- Bolker, B. (2014), 'Package lme4,' Accessed August 27, 2021. <http://cran.r-project.org/web/packages/lme4/lme4.pdf>.
- Boschma, R. (2005), 'Proximity and innovation: a critical assessment,' *Regional Studies*, 39(1), 61–74.
- Boschma, R. (2018), 'A concise history of the knowledge base literature: challenging questions for future research,' in A. Isaksen, R. Martin and M. Trippel (eds), *New Avenues for Regional Innovation Systems - Theoretical Advances, Empirical Cases and Policy Lessons*. Springer International Publishing: Cham, pp. 23–40.
- Boschma, R. and P.-A. Balland (2020), *Complementary Inter-Regional Linkages and Smart Specialization: An Empirical Study on European Regions*. Department of Human Geography and Spatial Planning, Group Economic Geography, Utrecht University: Utrecht.
- Boschma, R. and R. Martin (eds), (2010), *Handbook on Evolutionary Economic Geography*. Edward Elgar: Cheltenham.
- Breschi, S. and C. Catalini (2010), 'Tracing the links between science and technology: an exploratory analysis of scientists' and inventors' networks,' *Research Policy*, 39(1), 14–26.
- Breschi, S. and F. Lissoni (2009), 'Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows,' *Journal of Economic Geography*, 9(4), 439–468.
- Breschi, S., F. Malerba and L. Orsenigo (2000), 'Technological regimes and Schumpeterian patterns of innovation,' *The Economic Journal*, 110(463), 388–410.
- Broekel, T. (2015), 'The co-evolution of proximities – a network level study,' *Regional Studies*, 49(6), 921–935.
- (2019), 'Using structural diversity to measure the complexity of technologies,' *PLoS One*, 14(5), e0216856.
- Broekel, T. and R. Boschma (2012), 'Knowledge networks in the Dutch aviation industry. The proximity paradox,' *Journal of Economic Geography*, 12, 409–433.
- Broekel, T., P.-A. Balland, M. Burger and F. van Oort (2014), 'Modeling knowledge networks in economic geography: a discussion of four methods,' *The Annals of Regional Science*, 53(2), 423–452.
- Broekel, T. and M. Bednarz (2019), 'Disentangling link formation and dissolution in spatial networks: an application of a two-mode STERGM to a project-based R&D network in the German Biotechnology Industry,' *Networks and Spatial Economics*, 46, 677–704.
- Cantner, U., E. Conti and A. Meder (2010), 'Networks and innovation: the role of social assets in explaining firms' innovative capacity,' *European Planning Studies*, 18(12), 1937–1956.
- Capone, F. and L. Lazzarretti (2018), 'The different roles of proximity in multiple informal network relationships: evidence from the cluster of high technology applied to cultural goods in Tuscany,' *Industry & Innovation*, 25(9), 897–917.
- Carlino, G. and W. R. Kerr (2015), 'Agglomeration and innovation,' in G. Duranton, J. V. Henderson and W. C. Strange (eds), *Handbook of Regional and Urban Economics*. Vol. 5, Elsevier: North Holland, pp. 349–404.
- Castellacci, F. (2008), 'Technological paradigms, regimes and trajectories: manufacturing and service industries in a new taxonomy of sectoral patterns of innovation,' *Research Policy*, 37(6-7), 978–994.

- Coad, A. (2019), 'Persistent heterogeneity of R&D intensities within sectors: evidence and policy implications,' *Research Policy*, 48(1), 37–50.
- Cohen, W. M. and D. A. Levinthal (1990), 'Absorptive capacity: a new perspective on learning and innovation,' *Administrative Science Quarterly*, 35(1), 128.
- Coleman, J. S. (1988), 'Social capital in the creation of human capital,' *American Journal of Sociology*, 94, 95–120.
- Cowan, R., N. Jonard and M. Özman (2004), 'Knowledge dynamics in a network industry,' *Technological Forecasting and Social Change*, 71(5), 469–484.
- DeStefano, D. and S. Zaccarin (2013), 'Modelling multiple interactions in science and technology networks,' *Industry & Innovation*, 20(3), 221–240.
- Dosi, G., L. Marengo and C. Pasquali (2006), 'How much should society fuel the greed of innovators?' *Research Policy*, 35(8), 1110–1121.
- Dosi, G. and R. R. Nelson (2010), 'Technical change and industrial dynamics as evolutionary processes,' in B. Hall and N. Rosenberg (eds), *Handbook of the Economics of Innovation*, Vol. 1, Elsevier: North Holland, pp. 51–127.
- Edquist, C. and B. Johnson (1997), 'Institutions and organizations in systems of innovation,' in C. Edquist (ed.), *Systems of Innovation: Technologies, Institutions and Organizations*. Pinter Publishers: London, pp. 41–63.
- Emmert-Streib, F. and M. Dehmer (2012), 'Exploring statistical and population aspects of network complexity,' *PLoS One*, 7(5), e34523.
- Engen, O. A., E. O. Simensen and T. Thune (2018), 'The evolving sectoral innovation system for upstream oil and gas in Norway,' *Petroleum Industry Transformations*, 23–39.
- Erdős, P. and A. Rényi (1960), 'On the evolution of random graphs,' *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17–61.
- Fagerberg, J., D. Mowery and B. Verspagen (eds.) (2009), *Innovation, Path Dependency, and Policy*. Oxford University Press: Oxford.
- Fleming, L. and O. Sorenson (2001), 'Technology as a complex adaptive system: evidence from patent data,' *Research Policy*, 30(7), 1019–1039.
- Florida, R. (2005), 'The world is spiky globalization has changed the economic playing field, but hasn't leveled it,' *R Florida Atlantic Monthly*, 296(3), 48.
- Geels, F. W. (2004), 'From sectoral systems of innovation to socio-technical systems,' *Research Policy*, 33(6-7), 897–920.
- Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994), *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary societies/Michael Gibbons [and Five Others]*. SAGE Publications: London, Thousand Oaks, CA.
- Gill, J. and R. J. Butler (2003), 'Managing instability in cross-cultural alliances,' *Long Range Planning*, 36(6), 543–563.
- Giuliani, E. (2011), 'Role of technological gatekeepers in the growth of industrial clusters: evidence from Chile,' *Regional Studies*, 45(10), 1329–1348.
- Giuliani, E. (2013), 'Network dynamics in regional clusters: evidence from Chile,' *Research Policy*, 42(8), 1406–1419.
- Giuliani, E., P.-A. Balland and A. Matta (2018), 'Straining but not thriving: understanding network dynamics in underperforming industrial clusters,' *Journal of Economic Geography*, 30, 147–172.
- Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness,' *American Journal of Sociology*, 91(3), 481–510.
- Greunz, L. (2003), 'Geographically and technologically mediated knowledge spillovers between European regions,' *The Annals of Regional Science*, 37(4), 657–680.
- Gulbrandsen, M. and L. Nerdrum (2009), 'Public sector research and industrial innovation in Norway: a historical perspective,' in *Innovation, Path Dependency, and Policy*. J. Fagerberg, D. Mowery and B. Verspagen (eds), Oxford University Press: Oxford, pp. 61–88.
- Hardeman, S., K. Frenken, Ö. Nomaler and A. L. J. Ter Wal (2015), 'Characterizing and comparing innovation systems by different "modes" of knowledge production: a proximity approach,' *Science & Public Policy*, 42(4), 530–548.
- Hasson, J. A. and J. Tinbergen (1966), 'Shaping the world economy: suggestions for an international economic policy,' *Revue Économique*, 76(301), 840.
- Head, K. and T. Mayer (2014), 'Gravity equations: workhorse, toolkit, and cookbook,' in G. Gopinath, E. Helpman and K. Rogoff (eds), *Handbook of International Economics*. Vol. 4, Elsevier: North Holland, pp. 131–195.
- Heinze, G. and M. Schemper (2002), 'A solution to the problem of separation in logistic regression,' *Statistics in Medicine*, 21(16), 2409–2419.
- Hersoug, B. (2015), 'The greening of Norwegian salmon production,' *Maritime Studies*, 14(1), 1–19.

- Hidalgo, C. A. (2015), *Why Information Grows: The Evolution of Order, from Atoms to Economies*. Penguin Books: London.
- Hidalgo, C. A. and R. Hausmann (2009), 'The building blocks of economic complexity,' *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10570–10575.
- Hoekman, J., K. Frenken and F. van Oort (2009), 'The geography of collaborative knowledge production in Europe,' *The Annals of Regional Science*, 43(3), 721–738.
- Isard, W. (1956), *Location and Space-Economy*. MIT Press: Cambridge.
- Jacobs, J. (1970), *The economy of cities*, Vol. 584. Vintage Books: New York.
- Jaffe, A. B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations,' *The Quarterly Journal of Economics*, 108(3), 577–598.
- Janssen, M. J. and M. Abbasiharofteh (2022), 'Boundary spanning R&D collaboration. Key enabling technologies and missions as alleviators of proximity effects?,' *Technological Forecasting and Social Change*, 180, 121689.
- Juhász, S. and B. Lengyel (2018), 'Creation and persistence of ties in cluster knowledge networks,' *Journal of Economic Geography*, 121(6), 1203–1226.
- King, G. and L. Zeng (2001), 'Logistic regression in rare events data,' *Political Analysis*, 9(2), 137–163.
- Kunnskapsdepartementet. (2011), 'Nasjonal strategi for bioteknologi, 2011,' Accessed October 21, 2021. <https://www.regjeringen.no/globalassets/upload/kd/vedlegg/kompetanse/bioteknologistrategi.pdf>.
- Lata, R., T. Scherngell and T. Brenner (2015), 'Integration processes in European research and development: a comparative spatial interaction approach using project based research and development networks, co-patent networks and co-publication networks,' *Geographical Analysis*, 47(4), 349–375.
- Lazer, D. and A. Friedman (2016), 'The network structure of exploration and exploitation,' *Administrative Science Quarterly*, 52(4), 667–694.
- Lazzeretti, L. and F. Capone (2016), 'How proximity matters in innovation networks dynamics along the cluster evolution. A study of the high technology applied to cultural goods,' *Journal of Business Research*, 69(12), 5855–5865.
- Mäkitie, T., A. D. Andersen J. Hanson, H. E. Normann, and T. M. Thune (2018), 'Established sectors expediting clean technology industries? The Norwegian oil and gas sector's influence on offshore wind power,' *Journal of Cleaner Production*, 177, 813–823.
- Malerba, F. (2002), 'Sectoral systems of innovation and production,' *Research Policy*, 31(2), 247–264.
- Malerba, F. (2005), 'Sectoral systems of innovation: a framework for linking innovation to the knowledge base, structure and dynamics of sectors,' *Economics of Innovation and New Technology*, 14(1-2), 63–82.
- Malerba, F. (ed.), (2009), *Sectoral Systems of Innovation: Concepts, Issues and Analyses of Six Major Sectors in Europe*. Cambridge University Press: Cambridge.
- Malerba, F. and P. Adams (2013), 'Sectoral systems of innovation,' in *The Oxford Handbook of Innovation Management*. M. Dodgson, D. M. Gann and N. Phillips (eds), Oxford University Press: Oxford, pp. 183–203.
- Manniche, J., J. Moodysson and S. Testa (2017), 'Combinatorial knowledge bases: an integrative and dynamic approach to innovation studies,' *Economic Geography*, 93(5), 480–499.
- Marshall, A. (1920), *Principles of Economics*. MacMillan: London.
- Mazzucato, M. (2018), 'Mission-oriented innovation policies: challenges and opportunities,' *Industrial and Corporate Change*, 27(5), 803–815.
- Menon. (2017), 'Maritim verdiskaping,' *Maritim Forum*. Accessed October 21, 2021. <https://www.menon.no/wp-content/uploads/2017-Maritim-verdiskapingsbok.pdf> Report, retrieved.
- Micek, G. (2018), *Attempt at Summarising past Studies on Geographic Proximity*. Wydawnictwo Naukowe Uniwersytetu Pedagogicznego w Krakowie: kraków.
- Molina-Morales, X., J. A. Belso-Martínez, F. Más-Verdú and L. Martínez-Cháfer (2015), 'Formation and dissolution of inter-firm linkages in lengthy and stable networks in clusters,' *Journal of Business Research*, 68(7), 1557–1562.
- Morgan, K. (2004), 'The exaggerated death of geography: learning, proximity and territorial innovation systems,' *Journal of Economic Geography*, 4(1), 3–21.
- Nooteboom, B. (2000), 'Learning by interaction: absorptive capacity, cognitive distance and governance,' *Journal of Management and Governance*, 4(1/2), 69–92.
- Norsk Petroleum. (2018), 'EKSPORT AV OLJE OG GASS,' Accessed July 06, 2019. <https://www.norskipetroleum.no/produksjon-og-eksport/eksport-av-olje-og-gass/>.
- OECD. (2008), 'OECD science,' *Technology and Industry Working Papers 2008/02*.
- Owen-Smith, J. and W. W. Powell (2004), 'Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community,' *Organization Science*, 15(1), 5–21.
- Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory,' *Research Policy*, 13(6), 343–373.
- Petralia, S., P.-A. Balland and D. Rigby (2016), 'HistPat dataset: Harvard dataverse.'



- Ponds, R., F. van Oort and K. Frenken (2007), 'The geographical and institutional proximity of research collaboration,' *Papers in Regional Science*, 86(3), 423–443.
- Porter, M. E. (1998), 'Clusters and competition: New agendas for companies, governments, and institutions,' in M. E. Porter (ed.), *On Competition*. Harvard Business School Press: Boston, pp. 197–299.
- Powell, W. W., K. W. Koput and L. Smith-Doerr (1996), 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology,' *Administrative Science Quarterly*, 41(1), 116–145.
- Prehn, S., B. Brümmer and T. Glauben (2015), 'Gravity model estimation: fixed effects vs. random intercept Poisson pseudo-maximum likelihood,' *Applied Economics Letters*, 23(11), 761–764.
- Prestmo, J., B. Strøm and H. K. Midsem (2015), 'Ringvirkninger av petroleumsnæringen i norsk økonomi. Vol. 2015/8,' Accessed July 06, 2019. [https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/\\_attachment/352126?\\_ts=163cf7b62c0](https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/_attachment/352126?_ts=163cf7b62c0).
- Puhr, R., G. Heinze, M. Nold, L. Lusa and A. Geroldinger (2017), 'Firth's logistic regression with rare events: accurate effect estimates and predictions?,' *Statistics in Medicine*, 36(14), 2302–2317.
- Regjeringen. (2018), 'Long-term plan for research and higher education,' Accessed August 10, 2017. <https://www.regjeringen.no/en/topics/research/innsiktsartikler/langtidsplan-for-forskning-og-hogare-utdanning/id2353317>.
- The Research Council of Norway. (2016), 'Forskningsrådet samler de aller beste i et nytt nasjonalt senter,' Accessed November 10, 2016. [https://www.forskningsradet.no/no/Nyheter/Forskningsradet\\_samler\\_de\\_aller\\_beste\\_i\\_et\\_nytt\\_nasjonalt\\_senter/1254009252426?lang=no](https://www.forskningsradet.no/no/Nyheter/Forskningsradet_samler_de_aller_beste_i_et_nytt_nasjonalt_senter/1254009252426?lang=no).
- . (2018), 'The role of the Research Council,' Accessed September 26, 2018. [https://www.forskningsradet.no/en/The\\_role\\_of\\_the\\_Research\\_Council/1138785841810](https://www.forskningsradet.no/en/The_role_of_the_Research_Council/1138785841810).
- Reve, T. and A. Sasson (2012), *Et Kunnskapsbasert Norge*. Universitetsforl: Oslo.
- Rosiello, A. and A. Maleki (2021), 'A dynamic multi-sector analysis of technological catch-up: the impact of technology cycle times, knowledge base complexity and variety,' *Research Policy*, 50(3), 104194.
- Sarpebakken, B. and S. S. Ubisch (2017), *Ressursinnsatsen Til Marin FoU Og Havbruksforskning I 2015*. Nordisk institutt for studier av innovasjon, forskning og utdanning NIFU: Oslo.
- Schumpeter, J. A. (1911), *Theorie der Wirtschaftlichen Entwicklung (Theory of Economic Development)*. Duncker und Humblot: Berlin.
- Singh, J. (2005), 'Collaborative networks as determinants of knowledge diffusion patterns,' *Management Science*, 51(5), 756–770.
- Sonn, J. W. and M. Storper (2008), 'The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975–1997,' *Environment & Planning A*, 40(5), 1020–1039.
- Sorenson, O., J. W. Rivkin and L. Fleming (2006), 'Complexity, networks and knowledge flow,' *Research Policy*, 35(7), 994–1017.
- Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli and L. Pietronero (2012), 'A new metrics for countries "fitness and products" complexity,' *Scientific Reports*, 2(1), 723.
- Ter Wal, A. (2014), 'The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure,' *Journal of Economic Geography*, 14(3), 589–620.
- Ter Wal, A. and R. Boschma (2011), 'Co-evolution of firms, industries and networks in space,' *Regional Studies*, 45(7), 919–933.
- Timbergen, J. (1962), *Shaping the World Economy*. Twentieth Century Fund: New York.
- Torre, A. and A. Rallet (2005), 'Proximity and localization,' *Regional Studies*, 39(1), 47–59.
- Tóth, G., S. Juhász, Z. Elekes and B. Lengyel (2021), 'Repeated collaboration of inventors across European regions,' *European Planning Studies*, 2, 1–21.
- Usai, S., E. Marrocu and R. Paci (2016), 'Networks, proximities, and interfirm knowledge exchanges,' *International Regional Science Review*, 40(4), 377–404.
- Uzzi, B. (1997), 'Social structure and competition in interfirm networks: the paradox of embeddedness,' *Administrative Science Quarterly*, 42(1), 35–67.
- van der Wouden, F. (2020), 'A history of collaboration in US invention: changing patterns of co-invention, complexity and geography,' *Industrial and Corporate Change*, 29(3), 599–619.
- van der Wouden, F. and D. L. Rigby (2019), 'Co-inventor networks and knowledge production in specialized and diversified cities,' *Papers in Regional Science*, 98(4), 1833–1853.
- Wanzenböck, I. and K. Frenken (2020), 'The subsidiarity principle in innovation policy for societal challenges,' *Global Transitions*, 2, 51–59.
- Wanzenböck, I., J. Wesseling, K. Frenken, M. Hekkert and M. Weber (2020), 'A framework for mission-oriented innovation policy: alternative pathways through the problem-solution space,' *Science and Public Policy*, 47, 474–489.
- Weitzman, M. L. (1998), 'Recombinant growth,' *The Quarterly Journal of Economics*, 113(2), 331–360.
- Wuyts, S., M. G. Colombo, S. Dutta and B. Nooteboom (2005), 'Empirical tests of optimal cognitive distance,' *Journal of Economic Behavior & Organization*, 58(2), 277–302.

## Appendix A

**Table A1.** Overview of the structural properties of each sector's network

Sector	Time window	Number of components	Share of the largest component	Network density	Smallworldness <sup>a</sup>	Scalefreeness
Biotech	2005	7	0.949	0.042	9.759	0.814
	2008	22	0.884	0.025	12.843	1.014
	2011	20	0.929	0.018	15.893	1.034
	2014	14	0.956	0.018	16.409	1.075
Oil and gas	2005	6	0.952	0.057	6.02	0.675
	2008	4	0.991	0.037	8.527	0.837
	2011	7	0.966	0.03	9.583	0.881
	2014	13	0.95	0.029	10.123	0.897
Marine	2005	12	0.947	0.027	11.94	0.935
	2008	16	0.959	0.015	15.853	1.07
	2011	25	0.941	0.012	20.035	1.136
	2014	22	0.955	0.013	20.251	1.121
Maritime	2005	4	0.924	0.069	7.129	0.645
	2008	7	0.917	0.033	10.472	0.969
	2011	21	0.893	0.021	14.059	0.941
	2014	15	0.911	0.024	14.899	0.883

<sup>a</sup>Smallworldness is calculated for the largest component of the network.

## Appendix B Sectors' complexity

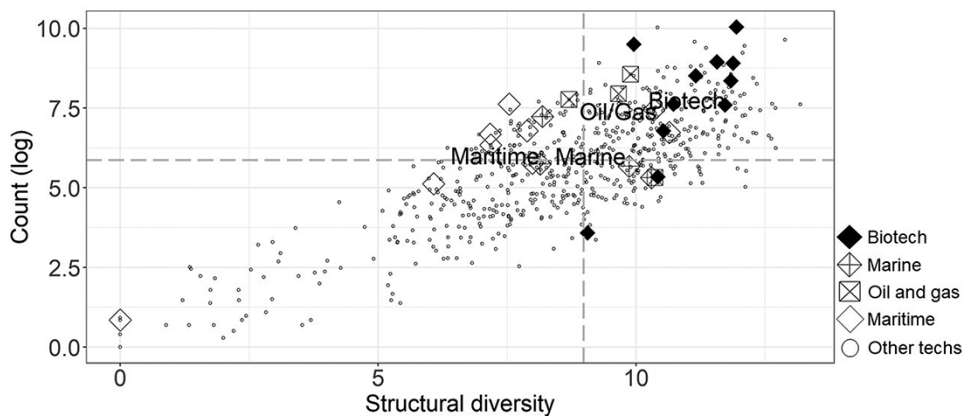
**Table B1.** List of IPC technological codes corresponding to each sector

Sector	Technological code	Description	
Biotech	A01H	New plants or processes for obtaining them; plant reproduction by tissue culture techniques	
	A61K	Preparations for medical, dental, or toilet purposes	
	C02F	Treatment of water, wastewater, sewage, or sludge	
	C07G	Compounds of unknown constitution	
	C07K	Peptides	
	C12M	Apparatus for enzymology or microbiology	
	C12N	Microorganisms or enzymes; compositions thereof; propagating, preserving, or maintaining microorganisms; mutation or genetic engineering; culture media	
	C12P	Fermentation or enzyme-using processes to synthesize a desired chemical compound or composition or to separate optical isomers from a racemic mixture	
	C12Q	Measuring or testing processes involving enzymes, nucleic acids, or microorganisms	
	G01N	Investigating or analyzing materials by determining their chemical or physical properties	
	C40B	Combinatorial chemistry; libraries, e.g., chemical libraries, in silico libraries	
	Marine	A01K	Animal husbandry; aviculture; apiculture; pisciculture; fishing; rearing or breeding animals, not otherwise provided for; new breeds of animals
		A22C	Processing meat, poultry, or fish
A23B		Preserving, e.g., by canning, meat, fish, eggs, fruits, vegetables, edible seeds; chemical ripening of fruits or vegetables; the preserved, ripened, or canned products	
A23J		Protein compositions for foodstuffs; working-up proteins for foodstuffs; phosphatide compositions for foodstuffs	

(continued)

**Table B1.** (Continued)

Sector	Technological code	Description
Oil and gas	C10B	Petroleum, gas, or coke industries; technical gases containing carbon monoxide; fuels; lubricants; peat
	B65D	Containers for storage or transport of articles or materials
	E21B	Earth or rock drilling; obtaining oil, gas, water, soluble or meltable materials, or a slurry of minerals from wells
	C09K	Materials for applications not otherwise provided for; applications of materials not otherwise provided for
Maritime	E02C	Ship-lifting devices or mechanisms
	E02B	Hydraulic engineering
	B63B	Ships or other waterborne vessels; equipment for shipping
	B63C	Launching, hauling-out, or dry-docking of vessels; life-saving in water; equipment for dwelling or working under water; means for salvaging or searching for underwater objects
	B65G	Transport or storage devices, e.g., conveyors for loading or tipping, shop conveyor systems, or pneumatic tube conveyors
	E02F	Dredging; soil-shifting
	E04H	Buildings or like structures for particular purposes; swimming or splash baths or pools; masts; fencing; tents or canopies, in general

**Figure B1.** Structural complexity of the four sectors

### B.1 Structural diversity (NDS) and frequency of technological classes

Individual NDS is a function of four structural properties (e.g., the structure of a collaboration network). Formally, [Emmert-Streib and Dehmer \(2012\)](#) define individual NDS for a network  $G$  by the following equation:

$$d(G) = \frac{\alpha_{module} \gamma_{motif}}{\nu_{module} \nu_{\lambda}}$$

where  $\alpha_{module}$  denotes the number of modules in  $G$  divided by the number of vertices,  $\gamma_{motif}$  corresponds to the share of motifs of sizes three and four.  $\nu_{module}$  denotes the variance of modules' size divided by their average.  $\nu_{\lambda}$  is the variance of the vector  $\Lambda(L)$  divided by its mean.  $\Lambda(L)$  is a set of eigenvalues for the Laplacian matrix<sup>xi</sup>  $L$ . This individual measure needs to be extended to ensure that the estimated value is not driven by chance ([Broekel 2019](#)). Therefore, the NDS is calculated based on a population of networks ( $G_M$ ) with similar structural properties to  $G$ .

$$NDS(\{G^S | G_M\}) = \frac{1}{S} \sum_{G \in G_M} d(G)$$

**Table B2.** Technology codes with the highest and lowest complexity scores between 2000 and 2010

Time window	CPC number	Rank in complexity	CPC	Description
2000–2010	643	1	B60L	Electric equipment or propulsion of electrically propelled vehicles; magnetic suspension or levitation for vehicles; electrodynamic brake systems for vehicles, in general
		2	C40B	Combinatorial chemistry; libraries, e.g., chemical libraries, in silico libraries
		3	H01L	Semiconductor devices; electric solid state devices not otherwise provided for
		641	C09F	Natural resins; French polish; drying-oils; driers (siccatives); turpentine
		642	G06C	Digital computers in which all the computation is effected mechanically
		643	E02C	Ship-lifting devices or mechanisms

## Appendix C

Table C1. Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) GEO	1													
(2) COG	0	1												
(3) INST	-0.01	0.17	1											
(4) SAME_AGGLOM	0	-0.02	-0.03	1										
(5) DIF_AGGLOM	0	-0.02	-0.03	-0.18	1									
(6) ONE_AGGLOM	0	-0.02	-0.03	-0.1	-0.14	1								
(7) SIZE(origin)	0.01	-0.12	-0.21	0.04	0.04	0.02	1							
(8) SIZE(destination)	0	-0.09	-0.11	0.01	0	0	0.02	1						
(9) PUBLIC(origin)	0	-0.14	-0.39	0.03	0.01	0.01	0.02	0.28	1					
(10) PUBLIC(destination)	0.02	-0.16	-0.6	0.04	0.05	0.03	0.35	0.01	0.01	1				
(11) HYBRID(origin)	0	-0.04	-0.05	0.03	0.02	0.01	0.12	0	-0.01	0.07	1			
(12) HYBRID(destination)	0	-0.04	-0.07	0.02	0.02	0.01	0	0.16	0.17	0.01	0	1		
(13) AGGLOM(origin)	0.01	-0.05	-0.12	0.22	0.36	0.13	0.12	0	0	0.21	0.05	0	1	
(14) AGGLOM(destination)	0	-0.04	-0.03	0.23	0.36	0.12	0	0.01	0.09	0	0	0.08	0	1

## Appendix D

**Table D1.** Results of regression models with a stepwise introduction of variables

	Marine Dependent variable: a dummy variable (each organization pair with at least one collaborative tie in marine takes the value of one, otherwise it takes the value of zero)		
	(1)	(2)	(3)
GEO			-0.009 (0.010)
COG			0.480 <sup>***</sup> (0.009)
INST			0.200 <sup>***</sup> (0.023)
SAME_AGGLOM		0.690 <sup>***</sup> (0.034)	0.682 <sup>***</sup> (0.034)
DIF_AGGLOM		0.105 <sup>***</sup> (0.032)	0.067 <sup>*</sup> (0.032)
ONE_AGGLOM		-0.146 <sup>***</sup> (0.021)	-0.137 <sup>***</sup> (0.021)
SIZE_origin	0.414 <sup>***</sup> (0.011)	0.403 <sup>***</sup> (0.011)	0.429 <sup>***</sup> (0.011)
SIZE_destination	0.461 <sup>***</sup> (0.010)	0.456 <sup>***</sup> (0.010)	0.474 <sup>***</sup> (0.010)
PUBLIC_origin	0.053 (0.056)	0.031 (0.057)	0.207 <sup>***</sup> (0.056)
PUBLIC_destination	-0.416 <sup>***</sup> (0.066)	-0.435 <sup>***</sup> (0.067)	-0.238 <sup>***</sup> (0.068)
HYBRID_origin	0.539 <sup>***</sup> (0.041)	0.515 <sup>***</sup> (0.041)	0.620 <sup>***</sup> (0.041)
HYBRID_destination	0.308 <sup>***</sup> (0.040)	0.306 <sup>***</sup> (0.041)	0.388 <sup>***</sup> (0.041)
AGGLOM_origin	0.193 <sup>***</sup> (0.020)	0.053 <sup>*</sup> (0.024)	0.058 <sup>*</sup> (0.025)
AGGLOM_destination	0.204 <sup>***</sup> (0.020)	0.034 (0.025)	0.072 <sup>***</sup> (0.025)
Constant	-7.146 <sup>***</sup> (0.123)	-7.034 <sup>***</sup> (0.124)	-7.584 <sup>***</sup> (0.126)
Organization-type FE	Yes	Yes	Yes
Observations	1,556,730	1,556,730	1,556,730
Nagelkerke's R <sup>2</sup>	0.066	0.071	0.093
Log-likelihood	-60,034.920	-59,717.680	-58,329.030
Akaike Information Criterion	120,135.800	119,507.400	116,736.100

\* $P < 0.1$ .\*\* $P < 0.05$ .\*\*\* $P < 0.01$ .



Maritime			
Dependent variable: a dummy variable (each organization pair with at least one collaborative tie in maritime takes the value of one, otherwise it takes the value of zero)			
	(1)	(2)	(3)
GEO			-0.041 <sup>***</sup> (0.015)
COG			0.553 <sup>***</sup> (0.013)
INST			0.050 (0.036)
SAME_AGGLOM		0.167 <sup>***</sup> (0.058)	0.154 <sup>***</sup> (0.058)
DIF_AGGLOM		-0.372 <sup>***</sup> (0.055)	-0.418 <sup>***</sup> (0.055)
ONE_AGGLOM		-0.259 <sup>***</sup> (0.034)	-0.256 <sup>***</sup> (0.034)
SIZE_origin	0.343 <sup>***</sup> (0.016)	0.341 <sup>***</sup> (0.016)	0.371 <sup>***</sup> (0.016)
SIZE_destination	0.362 <sup>***</sup> (0.016)	0.361 <sup>***</sup> (0.016)	0.380 <sup>***</sup> (0.016)
PUBLIC_origin	0.614 <sup>***</sup> (0.101)	0.599 <sup>***</sup> (0.101)	0.725 <sup>***</sup> (0.100)
PUBLIC_destination	0.045 (0.115)	0.032 (0.115)	0.179 (0.116)
HYBRID_origin	0.159 <sup>*</sup> (0.071)	0.155 <sup>*</sup> (0.071)	0.270 <sup>***</sup> (0.072)
HYBRID_destination	-0.084 (0.075)	-0.090 (0.075)	-0.002 (0.075)
AGGLOM_origin	0.095 <sup>***</sup> (0.031)	0.192 <sup>***</sup> (0.039)	0.207 <sup>***</sup> (0.039)
AGGLOM_destination	0.075 <sup>*</sup> (0.030)	0.162 <sup>***</sup> (0.039)	0.209 <sup>***</sup> (0.039)
Constant	-8.564 <sup>***</sup> (0.207)	-8.473 <sup>***</sup> (0.207)	-8.953 <sup>***</sup> (0.208)
Organization-type FE	Yes	Yes	Yes
Observations	1,556,730	1,556,730	1,556,730
Nagelkerke's $R^2$	0.039	0.042	0.068
Log-likelihood	-30,282.470	-30,194.640	-29,393.960
Akaike information criterion.	60,630.930	60,461.280	58,865.930

\*  $P < 0.1$ .\*\*  $P < 0.05$ .\*\*\*  $P < 0.01$ .

<b>Biotech</b>			
Dependent variable: a dummy variable (each organization pair with at least one collaborative tie in biotech takes the value of one, otherwise it takes the value of zero)			
	(1)	(2)	(3)
GEO			-0.007 (0.015)
COG			0.268*** (0.015)
INST			0.110*** (0.032)
SAME_AGGLOM		0.244*** (0.065)	0.246*** (0.065)
DIF_AGGLOM		-0.687*** (0.066)	-0.694*** (0.066)
ONE_AGGLOM		-0.437*** (0.039)	-0.429*** (0.039)
SIZE_origin	0.292*** (0.019)	0.286*** (0.019)	0.303*** (0.019)
SIZE_destination	0.285*** (0.017)	0.280*** (0.017)	0.291*** (0.017)
PUBLIC_origin	-0.275*** (0.100)	-0.326*** (0.101)	-0.218** (0.100)
PUBLIC_destination	-0.773*** (0.111)	-0.805*** (0.111)	-0.712*** (0.111)
HYBRID_origin	0.517*** (0.077)	0.516*** (0.078)	0.569*** (0.078)
HYBRID_destination	0.061 (0.077)	0.063 (0.078)	0.098 (0.078)
AGGLOM_origin	0.393*** (0.033)	0.510*** (0.044)	0.500*** (0.044)
AGGLOM_destination	0.436*** (0.032)	0.524*** (0.044)	0.531*** (0.044)
Constant	-6.044*** (0.185)	-5.806*** (0.186)	-6.096*** (0.187)
Organization-type FE	Yes	Yes	Yes
Observations	1,556,730	1,556,730	1,556,730
Nagelkerke's R <sup>2</sup>	0.079	0.088	0.094
Log-likelihood	-26,701.200	-26,431.930	-26,275.240
Akaike Information Criterion	53,468.400	52,935.860	52,628.490

\* $P < 0.1$ .\*\* $P < 0.05$ .\*\*\* $P < 0.01$ .

Oil and gas			
Dependent variable: a dummy variable (each organization pair with at least one collaborative tie in oil and gas takes the value of one, otherwise it takes the value of zero)			
	(1)	(2)	(3)
GEO			0.004 (0.016)
COG			0.405*** (0.014)
INST			0.152*** (0.037)
SAME_AGGLOM		1.669*** (0.048)	1.653*** (0.048)
DIF_AGGLOM		1.244*** (0.045)	1.198*** (0.046)
ONE_AGGLOM		-0.057* (0.032)	-0.045 (0.032)
SIZE_origin	0.704*** (0.020)	0.662*** (0.020)	0.679*** (0.020)
SIZE_destination	0.767*** (0.017)	0.749*** (0.017)	0.757*** (0.017)
PUBLIC_origin	0.583*** (0.082)	0.663*** (0.086)	0.747*** (0.085)
PUBLIC_destination	0.011 (0.084)	0.122 (0.092)	0.282*** (0.095)
HYBRID_origin	1.081*** (0.051)	0.998*** (0.052)	1.082*** (0.052)
HYBRID_destination	0.480*** (0.062)	0.484*** (0.062)	0.584*** (0.062)
AGGLOM_origin	0.344*** (0.033)	-0.252*** (0.037)	-0.236*** (0.038)
AGGLOM_destination	0.385*** (0.032)	-0.317*** (0.038)	-0.274*** (0.038)
Constant	-13.107*** (0.272)	-13.193*** (0.276)	-13.551*** (0.278)
Organization-type FE	Yes	Yes	Yes
Observations	1,556,730	1,556,730	1,556,730
Nagelkerke's $R^2$	0.118	0.139	0.153
Log-likelihood	-25,275.700	-24,671.260	-24,292.250
Akaike information criterion	50,617.410	49,414.530	48,662.510

\*  $P < 0.1$ .\*\*  $P < 0.05$ .\*\*\*  $P < 0.01$ .



## Appendix F

Table F1. Results of the regression models using Firth's bias reduction method (also known as the penalized likelihood method)

	Dependent variable: a dummy variable (each organization pair with at least one collaborative tie takes the value of one, otherwise it takes the value of zero)		
	Marine	Maritime	Biotech
COG	0.480*** (0.009)	0.477*** (0.010)	0.268*** (0.015)
SAME_AGGLOM	0.682*** (0.034)	0.553*** (0.013)	0.229*** (0.018)
COG×SAME_AGGLOM	0.014 (0.021)	0.155*** (0.058)	0.247*** (0.065)
GEO	-0.009 (0.01)	-0.073 (0.035)	0.151*** (0.065)
Constant	-7.581*** (0.125)	-0.041*** (0.015)	0.033 (0.015)
Controls	Yes	-8.944*** (0.208)	-6.094*** (0.186)
Organization-type FE	Yes	Yes	Yes
Observations	1,556,730	1,556,730	1,556,730
Likelihood ratio test	11,498.69***	4160.841***	5325.311***
Wald test	11,735.29***	4071.528***	6348.511***
			8621.756***
			8192.387***

\* $P < 0.1$ .\*\* $P < 0.05$ .\*\*\* $P < 0.01$ .