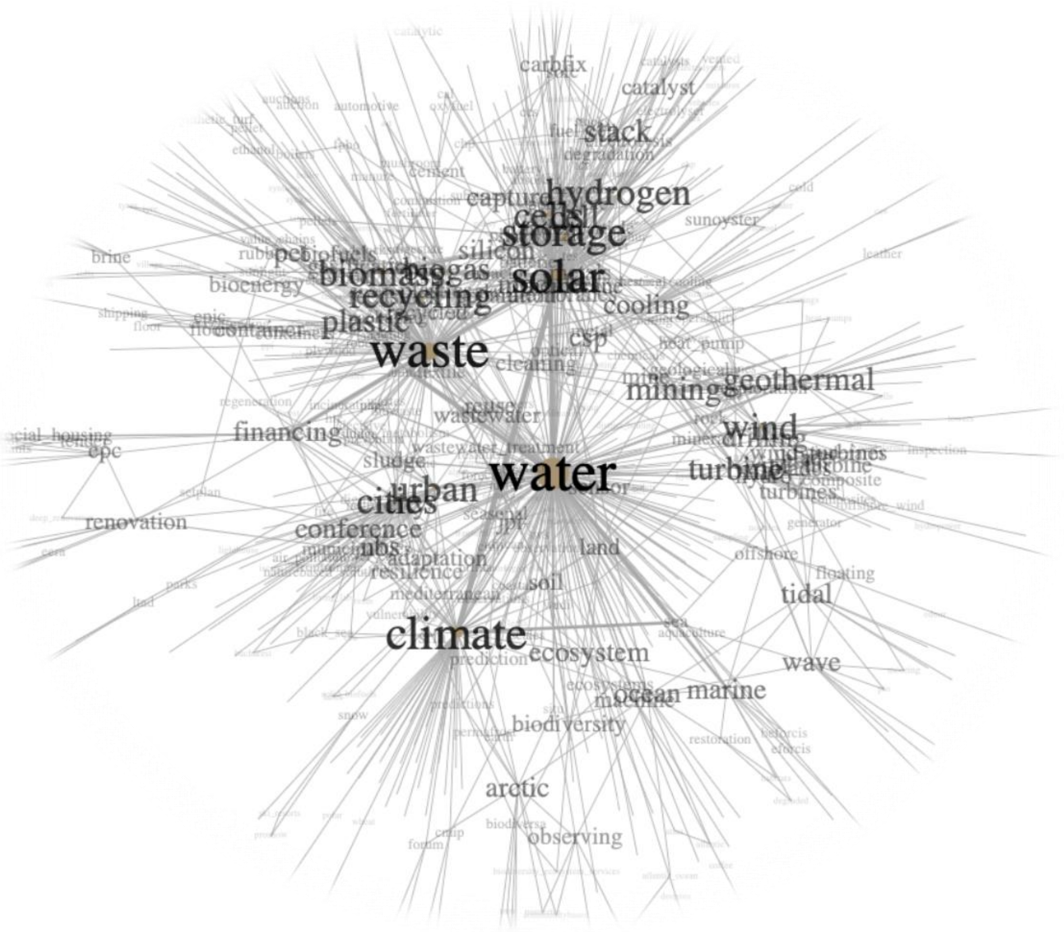
The equivocal role of proximity dimensions on technological novelty

A deep learning approach to exploring technological novelty in the Dutch energy sector



Copernicus Institute of Sustainable Development
Faculty of Geosciences
Utrecht University

Master Thesis Innovation Sciences

Name: Maurice Lekkerkerker
Student no.: 3873943
Date: 02-07-2023
Supervised by: dr. ir. Matthijs J. Janssen

Abstract

Technological development is driven by the recombination of related technological components and is occasionally disrupted through the introduction of uncommon combinations. These novel combinations can result in radically different technologies with profound economic and societal impact. However, the mechanisms through which technological novelty is created are not well understood. This complicates policy making in its aim to influence technological novelty. To address this research gap, this study measures the influence of proximities between R&D project partners on two types of technological novelty: (1) *structural* novelty, which involves the combination of disparate technological components and (2) *functional* novelty, which involves the introduction of technologies new to the system. We develop a text-based approach for quantifying technological novelty by utilizing GloVe word vectors obtained from energy R&D project abstracts. The novelty measures are validated and regressed against proximity characteristics of the R&D projects derived from Boschma's (2005) proximity framework. We found mixed results. Organizational proximity was found to have a negative effect on structural novelty for some technologies but a positive effect on functional novelty for energy efficiency projects. Technological and geographical proximity were found to be negatively associated with functional novelty for solar and wind energy technologies respectively. The remaining tested relationships were not significant. Despite these mixed results, our study highlights the value of using a text-based approach to measuring technological novelty. This offers a potentially valuable tool for policy makers in their efforts to ex ante detect technological novelty.

Acknowledgements

I would like to express my appreciation for my supervisor, Matthijs Janssen, for his guidance and support provided throughout the process of completing this master thesis. His accurate feedback, commitment and swift replies played a crucial role in shaping my research, without which completion would not have been possible. Moreover, I truly appreciate his openness towards my desire to experiment with novel text-based methodologies, which definitely enriched my overall research experience. Engaging in this process has strongly sparked my interest in NLP, in which I aspire to keep developing my expertise. I would also like to thank Maryse Chappin for her involvement and feedback on a preliminary version of this research. For providing the data used in my study, I would like to thank RVO and Joost Koch. I truly hope my research endeavors have been of added value

Furthermore, I would like to thank all individuals who have assisted me more indirectly. Els de Jong and Bram Kon in particular, who repeatedly expressed their encouragement during the process. Your support has been very valuable in pursuing me towards handing in the final thesis. Lastly, I would also like to thank my parents for their determined support and belief throughout this challenging yet educational journey. Their understanding, as well as participation in productive brainstorming, have been a source of motivation and I am grateful for their presence during the highs and lows of this academic endeavor. Their help and time have been instrumental in this achievement.

1. Introduction

Technological novelty is considered a pivotal source of economic development and societal change (Schumpeter, 1934; Nelson, 1993). Developments in biotechnology, artificial intelligence and communication, among other fields, have drastically changed the way we live and have significantly contributed to economic growth. Past advancements support our aspirations for addressing contemporary societal challenges to substantially rely on technological development. As such, understanding the drivers of technological novelty is essential for policy makers in their aim to reap potential societal benefits.

Despite alleged social and economic desirability, introducing novel technology is by no means self-evident. Technological development generally follows a rather predictable trajectory of incremental change. Only occasionally radically novel technologies are introduced that disrupt a prevailing paradigm (Dosi, 1982). The scarcity of technological discontinuity can be ascribed to the nature of technological search. During their search economic actors can benefit from knowledge previously accumulated in their technological domain (Nelson & Winter, 1982; Malerba, 1992), meaning that new R&D activities are generally closely related to prior search activities (Stuart & Podolny, 1996). This constrains technological advances towards incremental rather than radically novel technology, since radical inventions rely on inherently different combinations of knowledge and search strategies (Cohen & Levinthal, 1989; Ahuja & Lampert, 2001; Schoenmakers & Duysters, 2010). Furthermore, it is evident that not every introduction of technological novelty can result in equally successful radical invention. In fact, search activities involving more radical combinations of knowledge are known to generate less useful inventions on average (Fleming, 2001). Even if successful, the societal and economic impact might be realized only after a long period or even by other inventors drawing on the initial design. This makes novel technological search a risky and costly endeavor, subject to high uncertainty of outcomes and appropriability (Nelson, 1959; Arrow, 1962; Fleming, 2001).

Collaborative R&D projects during which actors jointly develop new technology can play an important role in this regard. In the course of technological trajectories, actors gain technological knowledge and acquire capabilities, constituting their present knowledge base and resources (Dosi, 1988; Cohen & Levinthal, 1990; Nieto & Santamaria, 2007). Collaborating allows actors to tap into each other's knowledge base, thereby expanding recombinatorial opportunities during technological search (Singh & Fleming, 2010). Consequently, the result of collaborative R&D projects can be novel combinations of knowledge reaching beyond the knowledge previously accumulated within a technological paradigm (Nelson & Winter, 1982). From a societal and economic perspective this is desirable, since radical innovations strongly rely on combining specialist diverse knowledge (Fleming, 2001; Schoenmakers and Duysters, 2010; Castaldi et al., 2015). Furthermore, collaborative R&D allows actors to confine the risks of highly uncertain novel technological search, by sharing resources and investments (Green et al., 1995).

Not every R&D project yields similar results in terms of innovative output. Finding novel combinations entails a successful conjunction of complementary bodies of knowledge and the learning of new organizational routines (Henderson & Clark, 1990; Cohendet & Llerena, 1997). Depending on the conditions in which knowledge exchange takes place, these knowledge combinations can be more or less novel. The degree of *proximity* between collaborating partners is generally considered an important factor in this regard. It affects uncertainties, coordination problems and knowledge exchange conditions, which in turn facilitate or hinder innovation and interactive learning (Gertler, 1995; Boschma, 2005; Nooteboom et al., 2007). Traditionally geographic proximity has received most attention in economic geography studies but more recently the importance of non-spatial proximity dimensions has been stressed (Knoben & Oerlemans, 2006). Boschma's (2005) work, which distinguishes between cognitive, organizational, social, institutional and geographical proximity, has been particularly influential. He argues that with regard to effective collaboration, an optimal degree of proximity between partners exists. If proximity is too low (i.e., actors differ too greatly), collaborating might not be fruitful as it is not possible to internalize the skills or competencies of the other (Hamel, 1991; Boschma & Lambooy, 1999). On the other side, too much proximity between actors makes collaborating redundant as no new knowledge can be obtained. This way, the output of collaborative R&D projects is subjected to influences of proximity between actors.

The influence of R&D partner proximity on project level technological novelty is not profoundly studied. One complicating factor is that technological novelty is typically measured ex post, after innovative output has become apparent in patent or product statistics. These statistics are often not available for pre-commercial R&D projects and as such many studies resort to different levels of analysis. On the firm level, strategic management literature is concerned with the influence of alliance portfolio proximity on individual firm performance (see: Branstetter & Sakakibara, 2002; Van Beers & Zand, 2014; Baum et al., 2000; Mouri et al., 2012; Miller, 2006). On the regional level, studies from economic geography focus on the relationship between proximities of economic actors and subsequent economic development of regions (Torre & Wallet, 2014). On a systemic level, studies from innovation systems literature emphasize the importance of R&D policy infrastructures and their implication for collaborating in novel technological search (Jacobsson & Bergek, 2004). However, none of these perspectives make explicit how the proximities in collaboration are related to project level technological novelty creation. As such, the relationship between proximities of R&D project partners and creation of technological novelty on the project level remains largely uncovered. This study aims to bridge that gap by answering the following research question:

How does proximity between project partners relate to the creation of technological novelty in collaborative R&D projects?

To answer the research question, we build hypotheses based on R&D partners' distances on different proximity dimensions and on two measures of novelty: (1) *structural* novelty, which involves the combination of disparate technological components and (2) *functional* novelty, which involves the introduction of technologies new to the system. Employing a text-based machine learning approach, these hypotheses are empirically tested on Dutch R&D projects that received funding under the "Topsector energy policy" between 2012 and 2016. Technological novelty is particularly salient in the energy sector in light of sustainability challenges, such as resource scarcity, growing energy demand and anthropogenic global warming (Kemp, 1994; Sagar & Van der Zwaan, 2006). These challenges are deeply coupled with path dependencies and lock-ins (Unruh, 2000; Markard & Truffer 2006) to which optimizing prevailing paradigms through incremental inventions is insufficient (Geels, 2002; Markard et al., 2012). Rather, it requires radically novel combinations of knowledge to introduce technological breakthroughs that enable shifts in technological trajectories (Nelson & Winter, 1982; Dosi, 1982; Nemet, 2009). Regarding Dutch energy policy, the aim is however not to solely promote breakthrough energy technology but also to retain the competitiveness of the Dutch energy sector, which can be realized through incremental innovation as well (Janssen et al., 2017).

The contribution of this study is threefold. First, we provide insight in the creation of technological novelty by R&D projects in the Dutch energy sector. Secondly, we add to the existing proximities framework, by empirically examining the proximity-related mechanisms that underlie technological novelty. These insights are relevant to innovation scholars and may further allow policy makers to better direct technological novelty through public R&D funding. Thirdly, we propose a machine learning approach to measuring technological novelty based on textual content. Thereby we contribute to existing literature on the operationalization of radical invention. This can be relevant for innovation scholars, policy makers and innovation managers as it allows for ex ante identification of potential radical innovation.

The remainder of this study proceeds as follows: In the theory section, technological novelty as well as other relevant theoretical concepts are conceptualized and hypotheses are formulated. Thereafter, in the methodology section these are operationalized. Then, the results section sets out descriptive statistics, after which the results for each hypothesis are discussed individually. Section 5 then discusses theoretical and societal implications of the study, after which final conclusions and an answer to the research question are formulated in section 6.

2. Theory and hypotheses

This section sets out the concept of technological novelty as a process of recombination of existing technological components and describes two ways of defining it. We then hypothesize what proximity related mechanisms are expected to underlie technological novelty creation.

2.1 Technological Novelty

Technological novelty is introduced when economic actors successfully conduct technological search for novel combinations of complementary but previously existing components of technological knowledge (Schumpeter, 1934; Basalla, 1980; Nelson & Winter, 1982; Fleming, 1999; Arthur, 2007). A component can be defined as a distinct physical portion of a technology, embodying a core design concept (Henderson & Clark, 1990). Examples are hydropower generators, consisting of a basin, turbines, a generator, and power transformers; or synthesizers consisting of oscillators, electronic filters and an amplifier. Every new combination that is invented adds a new element of technological knowledge to the existing universe of knowledge elements (Fleming and Sorensen, 2004; Ahuja & Lampert, 2001).

The degree to which inventions introduce technological novelty is typically defined in terms of their radicality. Empirical studies in evolutionary economic literature commonly consider an invention to be radical based on its impact on future technological advances. In this vein, radical inventions can form the basis of new technological paradigms on which future technological trajectories are built (Dosi, 1982). In management literature, an invention is denoted radical based on its subsequent impact on markets and industries (Anderson & Tushmann, 1990, Utterback 1996). Radical inventions open new markets, thereby disrupting existing modes of operation which requires industries to drastically adapt their routines and competences.

Defining technological novelty using impact related conceptualizations has a major drawback since it only takes into account those inventions that eventually developed into successful disrupting innovation. Due to high outcome variability (Fleming, 2001), radical invention measured ex post neglects those inventions that did not come to fruition, causing a research bias in understanding the conditions under which radical inventions are developed (Verhoeven et al., 2016). To overcome this problem, we conceptualize radical invention in terms of its underlying technological characteristics, considering its structural components and functionality. We follow the work of Dosi (1998) and Arthur (2007), who define technology as a combination of *components* that fulfill a specific need or *function* in society (e.g., cars and trains satisfying transportation needs or gas and wind turbines fulfilling the function of electricity generation). Thus, technologies differ from each other in the way they combine these building blocks to fulfill a specific need.

To this end, we distinguish between *structural* and *functional* technological novelty (Arthur, 2007; Verhoeven et al., 2016). When defining the degree of novelty, one can map existing technological components on a technology space, in which related components are located close to one another (Jaffe, 1986; Silverberg & Verspagen, 2005; Novelli, 2015). Components are closer related when previous recombination has been proven successful. Economic actors conduct searches for new inventions by exploring this space and recombining different technological components, thereby creating technological novelty (Fleming, 2001). Evaluating to what extent a combination introduces technological novelty implies comparing new artefacts to existing components in the technology space. *Structural* novelty is then defined as the degree to which technological combinations incorporate components located in distant parts of the technology space. This is a capability 'broadening' exercise, as it requires actors to combine previously unrelated knowledge domains (Argyres and Silverman, 2004). The concept of structural novelty is profoundly studied and considered to be an important driver of radical invention (Nooteboom, 2000; Fleming, 2001; Nemet, 2009).

Functional novelty entails novelty in functionality of a specific combination of technological components. It applies to a set of technologies and is therefore a macro-level concept (Paez-Aviles, 2017). It is defined as the degree to which new combinations are different compared to other designs within a system (Van Rijnsoever et al., 2015; Strumsky & Lobo, 2015). Van Rijnsoever et al. (2015) study a similar concept referred to as technological diversity, defined as the increase in available alternatives to fulfill a specific need. The concept is also empirically studied by Carnabuci & Operti (2013), who refer to it as recombinant reuse. Functional novelty increases recombinatorial opportunities in innovation systems and inhibits the potential for technological lock-in (Dosi, 1982; Van Den Bergh, 2008). On the other hand, a lack of functional novelty increases the chance that superior combinations remain unexplored. Furthermore, it decreases a system's technological resilience to unforeseen environmental fluctuations, particularly during the emerging phase of technologies (Negro et al., 2008). The introduction of functionally novel technologies can occur either through a novel combination that deviates from established practices or through the improvement of an existing combination fulfilling a specific function in a new way. The latter requires economic actors that improve technological combinations and find new application purposes, consequently 'deepening' their current capabilities (Argyres and Silverman, 2004). Examples of functionally novel technologies would be the use of hydroelectric dams for energy storage to solve intermittency problems of renewable energy generation or the use of gas pipe technologies for hydrogen transport.

Structural and functional novelty are two orthogonal concepts. Christensen (1997) asserts that functional novelty can have a disruptive impact by opening new markets, even though initially structural novelty is only incremental. However, high functional novelty does not necessarily imply low structural novelty. An R&D project can be both combining previously unrelated components and also fulfill a societal function in a new way. Figure 1 below further illustrates the orthogonality of the concepts and provides examples.

		Structural Novelty	
		Low	**High**
Functional novelty	**High**	Combinations of technological elements that have been explored before (related components), but which deviate from previously existing technologies that fulfil the same function. *Example:* *Concentrated solar power technology is well explored in the broader technology space but would diverge from existing practices when used to fulfill the function of energy production in in The Netherlands.*	Combinations of technological components that draw on knowledge from disparate regions of the technology space, and at the same time diverge from existing technologies within an innovation system. *Example:* *The use of gas pipe technology to transport and store hydrogen. Hydrogen and LNG technologies are still relatively unrelated and their combined use would deviate from existing practices in The Netherlands to fulfill the function of energy transport and storage.*
	Low	Combinations of technological components that are relatively familiar to the innovation system, drawing on knowledge from strongly related regions in the technology space. *Example:* *Projects concerning technologies to improve refinery processes of conventional energy in The Netherlands. These combinations are previously explored by Dutch actors and are therefore stronger related to existing technologies in The Netherlands.*	Combinations of technological components which are not unprecedented in the innovation system but draw on knowledge from disparate areas of the technology space. *Example:* *The combination of digesters and wastewater treatment technologies to provide district heating in The Netherlands. The combination is relatively new in the broader technology space but its contribution to novelty in The Netherlands might be limited as it is relatively well explored there.*

Figure 1: Technological novelty matrix with examples

2.2 The proximity framework

Engaging in collaborative R&D is often instrumental to joint knowledge development and organizational learning (Hamel, 1991; Teece & Pisano, 1994; Colombo, 2003). The proximities framework (Boschma, 2005) provides a useful tool for studying the factors influencing interorganizational R&D outcomes. Existing literature distinguishes between multiple, sometimes overlapping dimensions of proximity that play a role in interorganizational collaboration. Our proposed framework follows a seminal literature review by Knoben & Oerlemans (2006), who study ambiguity and overlap between conceptualizations of proximity. They conclude that three distinctive dimensions of proximity can be discerned with regard to inter-organizational collaboration: technological proximity, geographical proximity and organizational proximity.

2.2.1 Technological proximity

Technological proximity refers to the similarity between knowledge bases of participants in an inter-organizational collaboration (Knoben & Oerlemans, 2006). An organization's knowledge base underlies its technologies and innovative solutions (Dosi, 1988). Dosi (1988, p. 1126) defines it as the "set of information inputs, knowledge and capabilities that inventors draw on when looking for innovative solutions".

Innovation literature attributes great importance to the diversity of knowledge resources with regard to technological search (Nelson & Winter, 1982). Interorganizational R&D is seen as a means to increase the chances of innovative technological combinations (Prahalad and Hamel, 1997; Ahuja, 2000). Since innovation is a process of recombining existing pieces of knowledge, higher diversity of knowledge bases in an R&D project increases the number of combinations that can potentially be tried. Empirically, knowledge base diversity is found to foster the innovation process by enabling participants in R&D projects to recognize innovative opportunities and build novel associations (Cohen & Levinthal, 1990; Nieto & Santamaria, 2007). Vice versa, if participants in R&D projects are specialized in identical technological domains, they tend to write off exploration of technological combinations outside their existing repertoire as an unnecessary risk (Levinthal and March, 1993). As such, too much technological proximity decreases the number of technological configurations to be considered and thus confines the scope for recombination (Galunic & Rodan, 1998). Although higher diversity increases recombinative potential, a minimum degree of technological proximity between partners is required for participants to learn from the collaborative party and use external knowledge to commercial ends (Cohen & Levinthal, 1990; Boschma, 2005).

Previous studies highlight the importance of diverse knowledge for the creation of technological novelty. Hill & Rothaermel (2003) argue that knowledge required for radical technological invention tends to reside outside the boundaries of the organization's knowledge base. Chandy & Tellis (1998) and Christensen (2013) show that the degree to which external knowledge is new to the actor's knowledge base is deemed an important driver of radical invention. Furthermore, Neffke et al. (2011) show that combining unrelated knowledge through cross-industry linkages is likely to induce radical innovation during technological search. Given the recombinant nature of innovation, we expect that converging more diverse knowledge bases (i.e., lower technological proximity between project partners) is reflected in higher structural technological novelty of the innovation projects they collaborate in.

H1a: The technological proximity of R&D project partners is negatively associated with the project's structural novelty.

Introducing functional novelty likely requires knowledge from different technological disciplines. Paez-Aviles et al. (2017) indeed show that diversity of knowledge bases is positively associated with the creation of functional novelty. On the contrary, one can argue that technological search for functional novelty is hampered by low technological proximity as it is more difficult to exchange specialized knowledge when inventors are specialized in distant technological areas. However, Carnabuci & Operti (2013) find that diverse knowledge bases can enhance the ability to introduce both structural and functional novelty. Therefore, we hypothesize that technological proximity is negatively associated with a project's functional novelty.

H1b: The technological proximity of R&D project partners is negatively associated with the project's functional novelty.

2.2.2 Geographical proximity

Innovation is intimately bound up with tacit knowledge exchange, which is increasingly difficult to achieve at larger geographical distances (Cooke et al., 1997). Local knowledge networks allow for physical gatherings and interpersonal relationships which favours trust and interactive learning between project partners (Boschma, 2005). Strong interpersonal relationships induce knowledge spillovers that contribute to the diffusion of knowledge to neighbouring organizations in related industries (Cooke, 2008; Aghion & Jaravel, 2015). Empirically, Capaldo & Petruzzelli (2014) confirm this view, as they find a negative relationship between geographical distance and R&D project performance. Furthermore, Jaffe et al. (1993) find that patents tend to be cited in the same geographical area as where prior art was located. Consequently, knowledge is largely geographically bounded and differs substantially from region to region (Cooke 2001, Florida, 2005).

Despite contributing to local knowledge spillovers, geographical proximity also increases the potential for technological lock-in as geographically proximate actors tend to develop a shared knowledge context (Boschma, 2005). The shared context that exists among geographically highly proximate project partners could reinforce the belief in existing knowledge interlinkages, which would reduce the tendency of organizations to test different recombinations or new functionalities for existing technologies (Fleming, 2001). Regarding technological novelty, one can imagine that R&D projects subject to close geographical proximity might lack novelty as promising combinatorial opportunities outside their region might be overlooked (Capaldo & Petruzzelli, 2015). The likelihood of combining more disparate knowledge bases might indeed increase with lower geographical proximity and thereby yield more technological novelty in R&D projects. This was empirically substantiated by Bunduchi (2013) who finds that an over-reliance on geographically proximate partners leads to an emphasis on incremental innovation, hampering the ability to engage in radical innovation. Conversely, Phene et al. (2006), Sidhu et al. (2007) and Su (2022) find that combining knowledge from spatially distant locations creates an increased potential for novelty in R&D projects. However, too much spatial distance might hamper the accessibility of novel technological knowledge and interactive learning (Boschma, 2005). Indeed, Nan et al (2018) find that geographical proximity has a U-shaped relationship with technological recombination in R&D projects. They argue that excessive geographical distance between R&D project partners limits the accessibility of knowledge novelty and therefore impedes the recombination of technological knowledge. As such, we hypothesize that geographical proximity between project partners has an inverted u-shaped relationship with structural novelty.

H2a: The project partners' geographical proximity has an inverted U-shaped relationship with the project's structural novelty.

Geographical distance between project partners can potentially result in specialized and tacit knowledge spillovers from one region to the other (Meyer-Krahmer & Reger, 1999). Since functional novelty is measured relative to an innovation system, knowledge residing outside that system potentially adds to functional novelty. As this study analyzes energy projects in the context of the Dutch innovation system, knowledge from foreign project partners might add to the functional novelty. This is increasingly likely when geographical proximity between participants is low. Paes-Aviles et al. (2017) indeed find that geographical distance between R&D project partners is positively associated with technological diversity, which is conceptually similar to functional novelty. Similarly, Fitjar & Pose (2014) find that international cooperation appears as the main source of radical product and process innovation. We therefore hypothesize that higher geographical proximity is negatively related to an R&D project's functional novelty.

H2b: The project partners' geographical proximity is negatively associated with the project's functional novelty.

2.2.3 Organizational proximity

Organizational Proximity is known to suffer from conceptual ambiguity as it is studied at different levels of analyses. This study adopts the inter-organizational perspective as posed by Boschma (2005). In this light an organization is defined as a set of formal and informal rules and routines that depict behavior of members in particular professional situations (Rallet & Torre, 1999), referred to as organizational arrangement. Organizational proximity can therefore be understood as the degree to which two or more organizations differ regarding their organizational arrangement (Torre & Rallet, 2005; Knoben & Oerlemans, 2006). By sharing the same routines organizations can more easily interact which facilitates interactive learning (Boschma, 2005). Furthermore, proximate actors share innovation infrastructures such as labor markets for technically skilled personnel, which enhances knowledge integration (Mahmood & Mitchell, 2004; Capaldo & Petruzzeli, 2014). A typical example of inter-organizational proximity is actors belonging to the same organizational group (e.g., Knowledge Institutes (KIs), Governmental Organizations or Industry), as each organization type represents a unique mode of organizational arrangement (Kirat and Lung, 1999; Shaw and Gilly, 2000; Nan et al, 2018).

Differences in organizational arrangement have the potential to influence technological novelty of R&D projects as organization types each have a different search focus (Capaldo & Petruzelli, 2015). Regarding *structural novelty*, Van Beers & Zand (2014) find that organizational diversity leads to a higher variety of knowledge intake in R&D projects. Involving KIs in R&D projects provides teams with access to tacit knowledge (Cockburn & Henderson, 1998) as well as codified knowledge, enabling teams to build upon state-of-the-art research (Fabrizio, 2009). KIs direct technological search by providing R&D project teams with an improved understanding of the technological space in which they search for solutions for technical problems (Rosenberg, 1990; Fleming and Sorenson, 2004; Du et al., 2014). Indeed, empirical research shows that universities are important sources of knowledge and engage more actively in knowledge discussions during collaborative R&D, which enhances the extent to which new ideas are developed (Belderbos et al., 2004a; Talab et al., 2020). Regarding *functional novelty*, collaboration with KIs is generally more targeted at developing innovations with the potential to open up new markets (Tether, 2002). This was also highlighted by Belderbos et al. (2004b), who find collaboration with KIs to be an effective way of achieving innovation that opens up new market segments.

Other types of organizational arrangement may also contribute to technological novelty in collaborative R&D. SMEs enjoy the advantage of less bureaucracy and fewer hierarchical layers. More informal communication allows for faster decision-making processes with fewer barriers to eliminate radically novel directions of search (Nooteboom, 1994). Therefore, small firms tend to explore new technological spaces often ignored by larger firms (Almeida & Kogut, 1997). Consequently, involvement of SMEs in collaborative R&D potentially contributes to widening of the search scope. On the other hand, innovative search by industry incumbents is typically strongly routinized resulting in reluctance towards exploring new areas of the technology space

(Utterback, 1996; Chandy & Tellis, 2000). The innovative inertia of firms increases with age and size and directs innovative search along the beaten tracks (Hannan & Freeman, 1984, Christensen, 1997). Furthermore, incumbent firms are embedded in an established industry network to which it is harder to properly value new technological opportunities (Hill & Rothaermel, 2003). Therefore, large firms dominate innovative activities in well-established areas of the technology space (Almeida & Kogut, 1997). To overcome inertia incumbent firms may resort to SME collaboration to derive functionally novel technological outcomes that expand their markets (Jang et al., 2017).

Based on the above we hypothesize that collaborative R&D projects having organizationally diverse partners leads to (1) higher structural technological novelty through partners contributing their unique set of knowledge and skills; and (2) higher functional technological novelty through widening the search scope.

H3a: Organizational proximity is negatively associated with the project's structural novelty

H3b: Organizational proximity is negatively associated with the project's functional novelty

2.3 Conceptual Model

To summarize, the conceptual model depicted in Figure 2 visualizes the tested hypotheses. Each arrow represents a hypothesis, and its corresponding sign indicates the anticipated direction of the relationship.

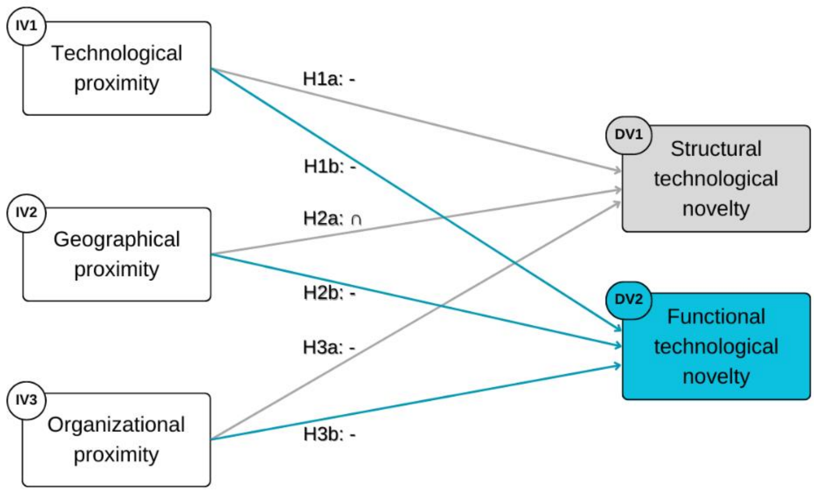


Figure 2: Conceptual model

3. Methodology

This section describes the data collection, preparation and analysis as well as the operationalization of the variables (Table 1) through which we test our hypotheses.

3.1 Research design

The aim of the present study is to empirically test the relationship between proximity among actors in collaborative R&D projects and the project's technological novelty. This relationship is quantitatively tested, taking collaborative energy R&D projects funded by the Dutch government as the unit of analysis. Thereby, this study adopts a quasi-experimental design in which the effects of non-randomly assigned independent variables on the dependent variable are estimated while controlling for confounding variables (Cook and Campbell, 1979).

3.2 Case description and data collection

Our data comprises 847 R&D projects that received funding under "Topsector energy policy" (TSE) policy between 2012 and 2016. Within TSE, funding is granted for different technology categories separately through tenders. The inclusion of technology categories is officially defined through 'programmalijnen' published in the Staatscourant[1]. Project data were obtained from innovation policy agency RVO and include the project abstracts, technology categories (IEA)[2] and participant information. Project data from energy policies preceding the TSE were also included. Project abstracts are subject to standardization set by RVO, requiring the project's motivation, description, objectives and results. In practice however, we find that not all project descriptions adhere to the proposed structure, resulting in a fairly high abstract length variety (Table 2). Any project descriptions in Dutch were translated to English using the Google Translate API. Any missing project abstracts were supplemented by data from the RVO website. Participant data were supplemented with industry classifications (SIC) obtained from the Chamber of Commerce. Out of all 847 R&D projects, 75 projects consist of only a single participant, which means that proximity statistics cannot be computed. These projects are therefore dropped from the analysis.

3.3 Operationalization of variables

Structural technological novelty (DV1)

Technological novelty is a relative measure by definition and can therefore only be measured in relation to what existed before. Hence, we define a technology space consisting of all existing technological artifacts to compare new R&D projects against. To construct this technology space, we propose a text-based approach that uses words and similarities between them, captured through word embeddings. Word embeddings are numerical vectors positioned in a vector space. They quantify the meaning of words in such a way that related words have similar positions in the vector space. This study uses pre-trained GlobalVectors (GloVe) which are 300 dimensional vectors learned from co-occurences of 840 billion words in publicly available texts up until 2014 (Peddington et al., 2014). Similar as to word combinations in texts, the extent to which knowledge elements are combined in technological artifacts specifies their relatedness (Ahuja & Lampert, 2001). In this vein, we use the similarity of words in project abstracts as the basis for our structural novelty measure.

---

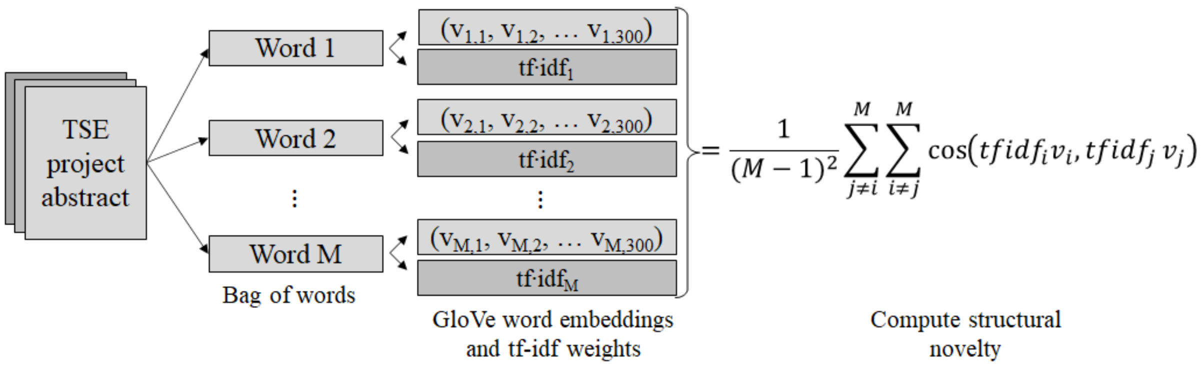
[1] A publication of the Dutch government in which generally binding regulations are published that have been established by ministerial regulation.

[2] International Energy Agency (IEA) distinguishes 24 subcategories of energy R&D. https://iea.blob.core.windows.net/assets/a2f370cf-873e-486f-935d-c2a117e14ba6/IEAGuidetoReportingEnergyRDDBudget-ExpenditureStatistics.pdf

When combining similar technology domains, the purpose of collaborative R&D is usually to obtain a deeper understanding of those domains, also referred to as specialization. These combinations are likely to have been already explored and it is expected these will likely lead to incremental innovation (Katila & Ahuja, 2002). In the technology space, these projects would draw from knowledge fields that are closely located to one another. When combining disparate pieces of knowledge, on the other hand, technologically deviant projects can lead to previously nonexistent artifacts. This is found relevant in two ways. Firstly, these projects broaden the scope of research (Katila & Ahuja, 2002). Secondly, combining knowledge from disparate domains potentially delivers more radical inventions (Schoenmakers & Duysters, 2010). In the technology space, these projects would draw from knowledge fields scattered over different parts of the space.

Structural technological novelty is computed by taking the average distance between pairwise combinations of TF-IDF weighted GloVe vectors of terms in a project's abstract after stop words are removed (Equation 1, below). By using TF-IDF weighted terms, we add more weight to words that stronger characterize that project relative to other abstracts in the corpus. Example abstracts with high and low structural novelty are reported in Appendix A and were found to be intuitive.

Equation 1



- M denotes the set of words
- v_i represents the index of word vector i in set M, v_j represents the index of vectors other than i
- $tfidf_i$ and $tfidf_j$ represent the tfidf weight of word vector i and j respectively
- The normalization factor $(M-1)^2$ avoids duplicate distance calculations for pairs (v_i, v_j)

To verify the reliability of this measure a robustness check is performed that tests whether structural technological novelty is correlated with technological (dis)continuity of a project. Discontinuity is defined as whether a new technology is strongly related to technologies currently part of the technological regime. This measure is part of our data from RVO. We find that technologically discontinuous projects are associated with higher technological novelty, which is in line with what one would expect (Figure 3). A t-test on the average differences of technological novelty between discontinuous and continuous projects confirmed that the differences were statistically significant $p < 0.001$.

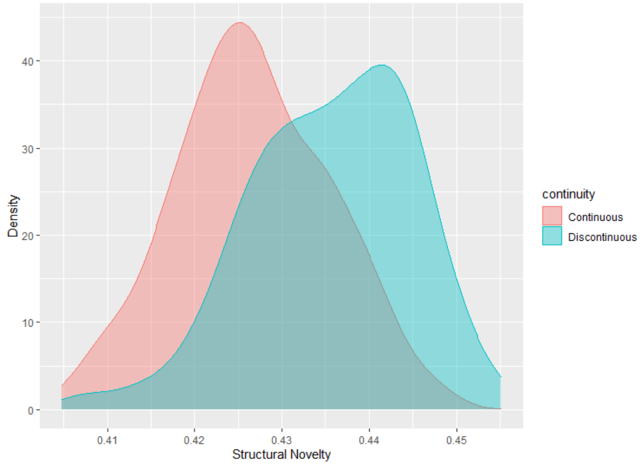
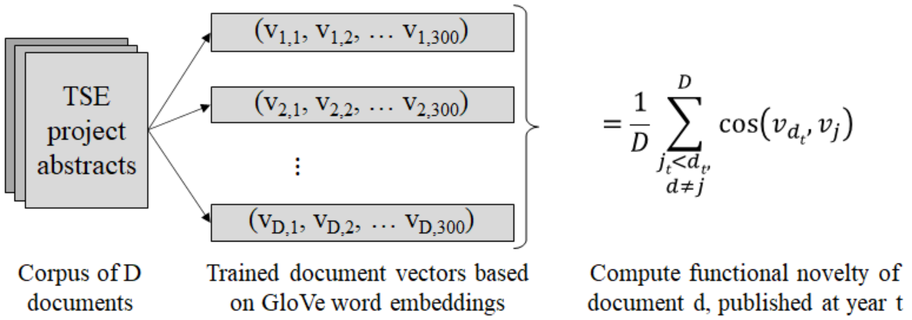


Figure 3: Density plot of structural technological novelty: continuous vs discontinuous projects.

Dependent variable 2: functional novelty

Since we define functional novelty as the degree to which recombinations are novel compared to existing technological combinations that fulfil a certain function, a textual representation of existing technologies in the Dutch innovation system is needed. To this end, we collect project abstracts from energy policies preceding the TSE. These include all projects funded under the Energie Onderzoek Subsidie (EOS, 2005-2010) and InnovatieAgenda Energie (IAE, 2008-2012). The functional novelty of project i is then computed by taking the average distance between the dense vector representation of document i compared to all preceding documents (Equation 2). These document vectors are training by applying the PV-DBOW algorithm on all project abstracts using the doc2vec package in R[3]. The training objective of the algorithm is to obtain document vectors that best predict the words within a document (Le & Mikolov, 2014). This way, it captures how documents differ from each other regarding their terminology. Just as for structural novelty, we tested whether a statistically significant difference between continuous and discontinuous projects existed for functional novelty, but such a relationship was not found. Section 4.2.2 further describes the validity of the functional novelty measure.

Equation 2



$$= \frac{1}{D} \sum_{\substack{j_t < d_t, \\ d \neq j}}^{D} \cos(v_{d_t}, v_j)$$

Compute functional novelty of document d, published at year t

- D denotes the set of documents
- v_{dt} represents the 300-dimensional document vector of focal document d, v_j represents document vectors preceding document d
- j_t<d_t indicates that all documents j precede d

[3] https://cran.r-project.org/web/packages/doc2vec/index.html

Independent variable 1: Technological proximity

Technological proximity is operationalized through a revealed relatedness measure between organizations that participate in collaborative R&D projects. The revealed relatedness concept was introduced by Neffke et al. (2016) who studied industry relatedness based on labour flows between 4-digit NACE codes. NACE presents a classification system used to categorize economic activities of organizations. Neffke et al. (2017) argue that the extent to which two industries hire workforces with similar knowledge and competences reveals their industry relatedness. In that vein a larger labour flow between industries indicates a higher proximity in terms of technological knowledge and vice versa. Our data describes pairwise industry labour flows in The Netherlands between 2009 and 2011 and was obtained via Utrecht University. The skill relatedness between industry i and j (SR_{ij}) is defined through Equation 3:

$$SR_{ij} = \frac{F_{ij}}{\sum_{k \neq j} F_{kj} \sum_{l \neq i} F_{il}} \sum_{F_{k(l \neq k)}} F_{kl} \qquad \textit{Equation 3}$$

Where F_{ij} is the total labour flow from industry i to j, F_{kj} is the total inflow of workers to industry j, F_{il} is the total outflow of workforce from industry i and F_{kl} represents the total movement of workers reported during the period of 2009 to 2011. Neffke et al. (2017) demonstrate the validity of the measure defined by equation 3 as it is stable over time and very similar across workers in different occupations and wage groups. The resulting relatedness values are then normalized to the [-1,1] interval to correct for positive skewness of the variable (Equation 4).

$$SR_{normalized} = \frac{SR_{ij} - 1}{SR_{ij} + 1} \qquad \textit{Equation 4}$$

The revealed relatedness between industries is asymmetrical (i.e., the skill relatedness of $i \rightarrow j$ is different from $j \rightarrow i$). To overcome this problem and reduce missing values, we symmetrize the measure by using the maximum relatedness value between industries *i* and *j*.

The use of solely a skill relatedness measure based on industry similarity might neglect the additional knowledge that two partners from identical industries might bring. To that end, we leverage the skill relatedness measure by accounting for the variety and balance of project partners (Stirling, 2007), which is computed through Equation 5.

$$SkillRelatedness_{weighted} = \sum_{ij(i \neq j)} SkillRelatedness_{ij} \cdot p_i \cdot p_j \qquad \textit{Equation 5}$$

Independent variable 2: Geographical proximity (IV2)

Using Bing Maps geocoder, the geocodes (longitude, latitude) of participants were retrieved based on the participant addresses provided by RVO. The average geographical distance between partners in an R&D project was computed using the R geosphere package[4], obtaining a value for the project's geographical proximity. The resulting average distance was then log transformed to account for the right skew of the variable originating from trans-continental partnerships (Broekel & Boschma, 2012).

[4] https://cran.r-project.org/web/packages/geosphere/geosphere.pdf

Independent variable 3: Organizational proximity / Organizational balance (IV3)
A typical operationalization of organizational proximity relies on categorizing organizations based on their institutional arrangements (Kirat and Lung, 1999; Shaw and Gilly, 2000). To this end, we distinguish different actor types within an R&D project; Large Enterprises (LE), Small and Medium Enterprises (SMEs) and Knowledge Institutes (KI). These actor types are defined in RVO project data. Organizations providing funding advisory services are excluded. To obtain group level organizational proximity we measure the Shannon Entropy index (H), a widely accepted measure to quantify group or team level diversity (Harrison & Klein, 2007). It describes randomness in a population. When partners in a project share one type of organizational arrangement, the group is considered highly organized as it is not characterized by a large degree of randomness (Shannon & Weaver, 1964), indicating high organizational proximity. Vice versa, different institutional arrangements jointly in an R&D project are less organized and exhibit higher randomness and lower proximity. The Shannon entropy is measured through Equation 6.

$$H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i log_2 p_i$$ *Equation 6*

where p_i denotes the proportion of organizational arrangement *i* in the project and *n* is the total number of organizational arrangements present in a collaborative R&D project. High entropy values indicate a more even distribution of project partners over organizational categories, implying a low degree of organizational proximity. Conversely, low entropy values denote a more uneven distribution of project partners over organizational categories, indicating higher proximity of organizational arrangements. For ease of interpretation, the organizational proximity variable is therefore renamed to *organizational balance* in the results section.

3.4 Control variables

Technology Readiness Level
Inherent to higher technological novelty is a longer time-to-market period. A systematic way to assess the time-to-market of a particular technology is the technology readiness level system (TRL), developed by NASA (Mankins, 2009). It comprises 9 levels, ranging from: scientific research without any orientation towards a priori specified objectives (level 1), to: R&D concerning actual application and commercialization of a technology, focused towards specific objectives (level 9). Since our data set consists of projects from all TRL, ranging from 1-9, one might expect that TSE projects concerned with lower TRL levels score higher on technological novelty. We therefore include TRL as a control variable.

Document length
The probability of sampling semantically different words increases with the number of unique terms in a document, which thus likely influences the technological novelty measure. When measuring technological novelty of a random sample of 25,000 US energy patents, we indeed find such a relationship (Figure 4, R = 0.11). The number of unique terms used in a document is therefore added as a control variable in our analyses.

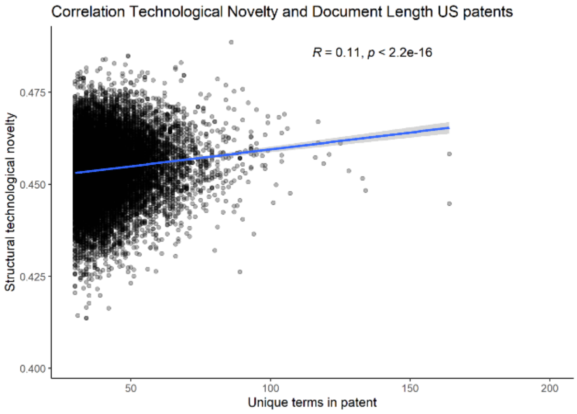


Figure 4: Correlation number of unique terms and technological novelty in US patents

Project starting year

Another variable that could affect the novelty measurement of documents is the starting year of a project. A reliable measure of technological novelty should detect novelty regardless of the year in which a project was started. Similarity between words might however depend on the year in which one measures. The similarity between the words 'battery' and 'vehicle' for example could have become stronger over time due to the technologies becoming more related. The embeddings used in this study are created from a large training set containing documents up until 2014[5] and as such could underestimate the novelty of older projects and overestimate the novelty of newer projects. Interestingly, we do not find such a relationship when computing structural novelty over the sample of US energy patents (Figure 5). Here our measure of technological novelty is similar for different patent filing years. Another way in which project starting year could influence the variables is through updates in regulation. Annually, the eligibility criteria for obtaining funding are revised in De Staatscourant based on progressive insights. This can impact the formation of R&D consortia. To exclude any temporal effects on our regression results, we include the project starting year as a control variable in our analysis.

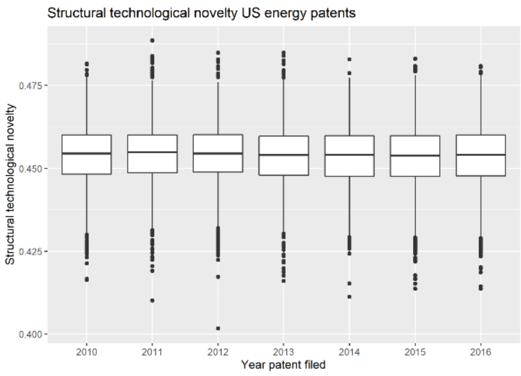


Figure 5: Technological novelty of a random sample of 25,000 US energy patents. Patents show a constant novelty distribution over time.

[5] https://nlp.stanford.edu/projects/glove/

Funding
The amount of funding a project receives potentially allows agents to search more extensively through the technology space which increases the probability of finding novel combinations. Previous research has found that larger grants are likely to increase the R&D scope of firms (Aschhoff, 2009). It must be noted however that funding is granted based on research proposals that are evaluated on 'innovativeness', meaning that the causality could also be the other way round, e.g., novel proposals potentially receive more funding. To this end, we analyzed 'innovativeness' scores assigned to proposals by industry experts during the proposal review phase. We indeed find significant positive correlations between proposal innovativeness and funding (Figure 6). Because of this existing relationship and the consensus in literature, we include funding as a control variable in our analysis.

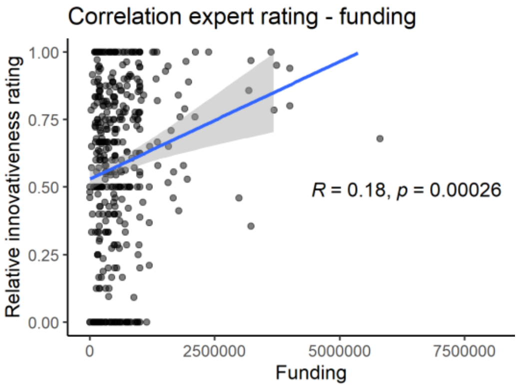


Figure 6: positive relationship between project funding and expert innovativeness rating

3.5 Data analysis
We started by removing any outliers from the data, based on the Cook's distance (Cook & Weisberg, 2012). Data analysis then proceeded in two steps. First, aggregated multiple linear regression models were fit to test whether variance in our dependent variables can be explained by the independent variables. We iteratively added variables after control variables and verified whether the models' Aikaike Information Coefficient (AIC) improved as a result. After the models were fit, we conducted further analysis to test whether the fitted models violated any of the multiple linear regression assumptions. For all models the variance inflation factors (VIFs) were calculated to determine whether our results were subject to multicollinearity (Cohen et al., 2003). Second, since funding is granted for different technology categories separately, we split data by IEA categories and repeated the analyses. This was done to uncover potentially masked relationships in the aggregated data.

Table 1: Operationalization of variables

Variable	Concept	Type
unique_terms	Unique terms in a project abstract	ordinal
funding	Amount of funding in € awarded by Dutch government	continuous
year	Project starting year	ordinal
trl_level	The project's technology readiness level (1-4)	ordinal
tech_proximity	Technological proximity between project partners based on similarity of SIC industry codes	[0,1]
log_distance_km	Log of geographical distance in km between project partners	Continuous
org_balance	Inverse of organizational proximity measured by the evenness of distribution over partner type categories in a project	Continuous [0,e]
structural novelty	Relatedness of technological artifacts in a project's abstract based on the similarity of tfidf weighted word vectors	Continuous [0,1]
functional novelty	The relatedness of a project's abstract to previous projects based on the similarity of document vectors	Continuous [0,1]

4. Results

The results section starts by providing insights into the data through descriptive statistics. Descriptive statistics on the structural and functional novelty variables will be discussed separately in 4.2.1 and 4.2.2 respectively. Then the results of the linear regression models are discussed, again separately per dependent variable (sections 4.3 and 4.4).

4.1 Descriptive Statistics

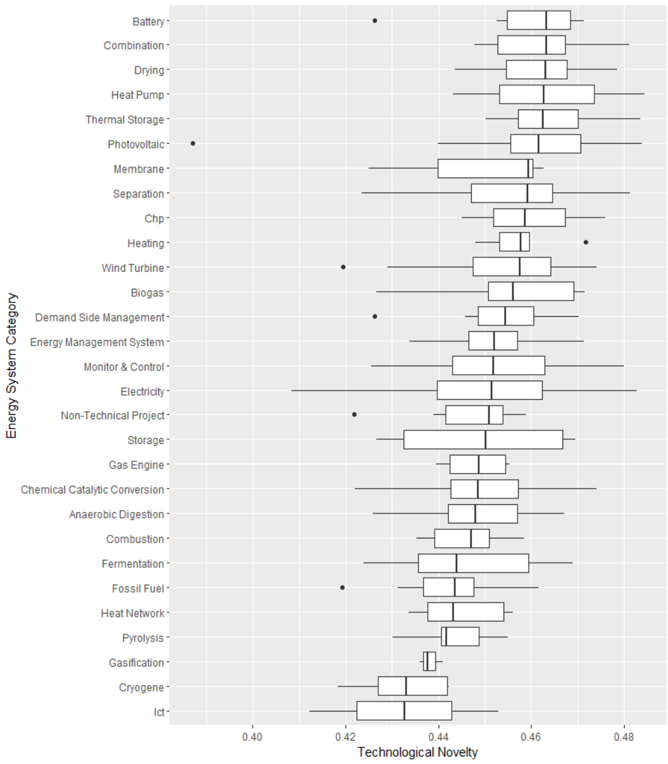
This section describes any correlations between control, independent and dependent variables. Since our statistical analyses are also applied on individual energy R&D technology categories, we also provide statistics on the IEA level. These can be found in Appendix E. Correlations among variables are presented in Table 2.

Table 2: Correlation table showing means, standard deviations, and correlations with confidence intervals

	Variable	*M*	*SD*	1	2	3	4	5	6	7
1	unique_terms	19.85	7.27							
2	funding	588963.70	891086.66	.08 [-.00, .16]						
3	n_participants	4.32	3.13	-.02 [-.11, .06]	.17** [.10, .24]					
4	tech_proximity	0.09	0.42	.07 [-.02, .15]	.05 [-.03, .13]	-.09* [-.16, -.01]				
5	log_distance_km	4.32	1.13	.02 [-.07, .10]	.07 [-.01, .15]	.20** [.12, .27]	.01 [-.06, .09]			
6	org_balance	0.57	0.43	-.07 [-.15, .02]	.11** [.04, .18]	.51** [.46, .57]	-.03 [-.10, .05]	.13** [.05, .21]		
7	structural_novelty	0.43	0.01	.44** [.37, .50]	.12** [.04, .20]	-.02 [-.11, .06]	-.01 [-.10, .07]	.02 [-.07, .11]	.05 [-.04, .13]	
8	functional_novelty	0.43	0.02	.21** [.13, .29]	.13** [.05, .20]	.14** [.07, .22]	.01 [-.07, .09]	.08 [-.00, .15]	.14** [.06, .21]	.11** [.03, .20]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. * indicates $p < .05$. ** indicates $p < .01$.

Regarding the number of unique terms in a document and technological novelty we observe a strong significant positive correlation (p<0.001), confirming its added value as control variable. It is intuitive to posit that the number of project partners (n_participants) is correlated with the independent variables, since a higher number of participants contributes to more diversity and less relatedness within a project team. The independent variables under study, however, are not significantly correlated with one another indicating the absence of any multicollinearity issues in the data.



system_category	structnov_variance
Storage	0.0004710
Separation	0.0002860
Battery	0.0002810
Fermentation	0.0002636
Electricity	0.0002586
Monitor & Control	0.0002405
Biogas	0.0002060
Chemical Catalytic Conversion	0.0001984
Non-Technical Project	0.0001842
Photovoltaic	0.0001820
Heat Pump	0.0001641
Wind Turbine	0.0001621
Combination	0.0001360
Anaerobic Digestion	0.0001346
Drying	0.0001330
Demand Side Management	0.0001247
Energy Management System	0.0001225
Fossil Fuel	0.0001143
Cryogene	0.0001032
Heat Network	0.0000977
Pyrolysis	0.0000652

Figure 7: Structural novelty (left) and Structural novelty variance (right) for different Energy System Categories

4.2.1 Descriptive statistics dependent variable: Structural Technological Novelty

The boxplot in figure 7 (left) shows structural novelty for energy projects grouped by their energy system categories, sorted from highest to lowest median novelty score. Energy projects concerning Batteries score highest on technological novelty, followed by Combination, Drying and Heat Pump projects. Interestingly, the Combination category which concerns projects that combine different energy system components ranks high on the technological novelty measure. This illustrates that our technological novelty measure seems to capture the recombinant nature of technology, contributing to the measure's validity. Lower ranked projects are more likely concerned with fossil fuels, such as Gas Engine, Combustion, Gasification, Pyrolysis and Cryogene (Liquified Natural Gas). Another remarkable observation is the difference in *variance* of technological novelty scores between system categories. Figure 7 (right) orders system categories from high to low variance, where low variance means categories with comparable technological novelty scores and vice versa. Overall, the lower variance categories seem to be associated with projects that involve more incumbent technologies, such as gas engine, pyrolysis and fossil fuels. On the other hand, higher variance seems to be associated with less technology specific categories such as Storage, (molecular) Separation and Electricity. This resembles the technological variety of for instance Storage projects as these might concern incumbent (gas & oil) as well as very novel energy carriers, such as hydrogen.

4.2.2 Descriptive statistics dependent variable: Functional Technological Novelty

The TSE abstracts' document vectors used to compute functional novelty are visualized in the t-sne plot (Maaten & Hinton, 2008) in figure 8. The visualization reveals some interesting insights regarding the way TSE projects are semantically related to one another. Basic (generic) energy research is logically located in the centre of the space, coinciding with its non-specific nature. Energy storage projects (dark blue) are widely dispersed across the space, which is intuitive as energy storage has applications in multiple areas (e.g., buildings, vehicles, energy production). Wind Energy projects are located next to Oil & Gas, suggesting that similar knowledge might be needed for these technologies. This possibly concerns offshore projects that require drilling and platform technologies. Mobility is positioned between Oil & Gas and Biofuels, which makes sense as both energy production categories have applications in the mobility sector (biogas, bio-ethanol). BioFuels is closely intertwined with industry energy efficiency. This resembles the programmatic intention of the Dutch government to stimulate bio-energy as an initial step towards industry sustainability.

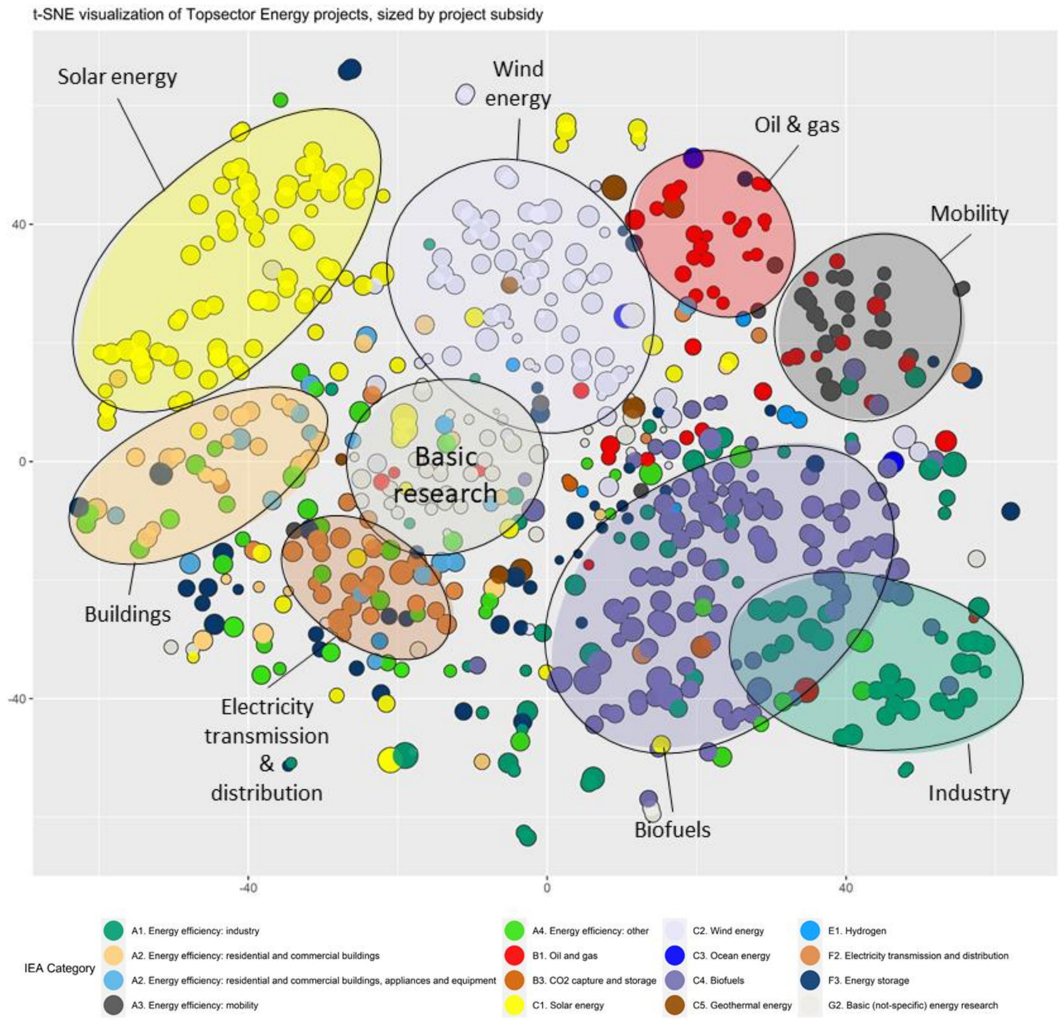


Figure 8: Document space visualizing the similarity between document vectors. Data points represent documents, coloured by their corresponding IEA category and size by functional novelty.

Taking a closer look at functional novelty we find that projects related to Energy Management Systems, Photovoltaic and Combination add the most novelty to the Dutch energy system on average (Figure 9). Gas technologies such as Cryogene (LNG storage) and Gasification add least novelty to the Dutch energy system. This is intuitive since The Netherlands historically has had a strongly developed gas sector and infrastructure. The ambition to increasingly replace gas technologies with electrical technologies such as PV or heat pumps seems reflected in this novelty score. When we categorize the projects into four buckets according to their functional novelty scores, an interesting pattern emerges. Projects associated with the terms 'fuel', 'businesses,' and 'encouraging' are predominantly found in the lowest functional novelty bucket (Figure 10, left), whereas 'alternatives,' 'changing,' and 'biogas' are predominantly present in the highest bucket (Figure 10, right). This juxtaposition of words, the former group indicating a shorter time-to-market and the latter group signifying a deviation from prevailing practices in The Netherlands, contributes to the validity of the functional novelty measure.

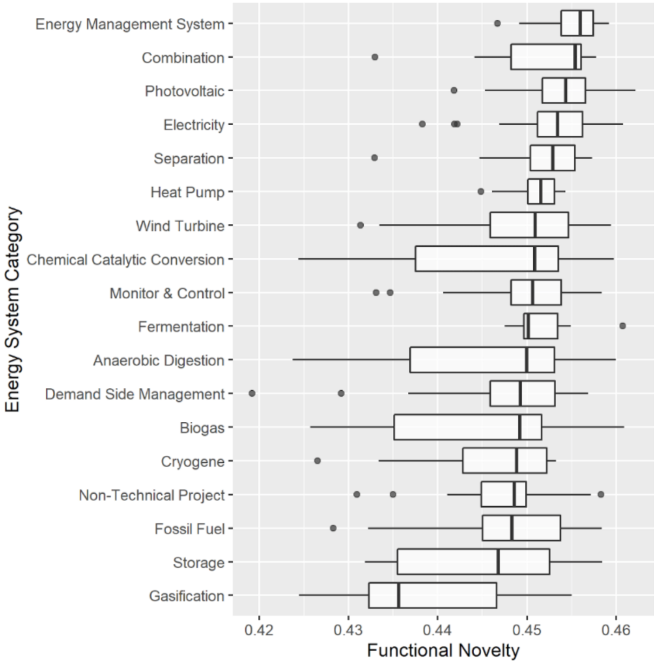


Figure 9: Functional novelty per energy system category

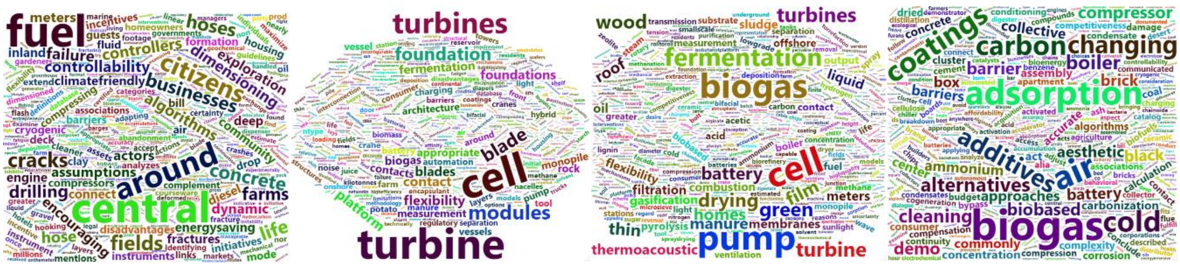


Figure 10: Word clouds functional novelty from lowest to highest functional novelty. Since some categories consist of more projects, words related to those projects are stronger visualized in the word clouds.

4.3 Regression results

This section discusses regression results for our first dependent variable: structural technological novelty. Structural novelty captures the degree to which words used in R&D abstracts are distant to one another in the vector space.

4.3.1 Regression results dependent variable I: structural novelty

Table 3 reports the regression results with structural novelty as dependent variable. We start by fitting a base model with just the control variables (Model 0). We find control variables unique_terms, trl_level_squared (p<0.001), year (p<0.01) and funding, trl_level (p<0.05) to be statistically significant in the base model. The number of unique terms is strongly significant for all models, which shows that our measure of structural novelty is strongly influenced by the unique number of words in a document. This is intuitive since having more unique words increases the likelihood of an abstract containing words in distant parts of the vector space. We continue by adding the hypothesized predictor variables while verifying whether this leads to an improvement of the model. After adding the predictor variables (Model 1, 2 and 3) we find that the Akaike Information Criterion (AIC) increases compared to the base model (Model 0), indicating that adding predictor variables does not improve the model fit. We therefore find no evidence supporting our first set of hypotheses (1A, 2A, 3A).

Table 3: Regression results using structural novelty as dependent variable

Predictor	Model 0	Model 1	Model 2	Model 3
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	**-1.357* (6.130e-01)**	-8.154e-01 (6.319e-01)	-9.531e-01 (6.254e-01)	-9.827e-01 (6.283e-01)
unique_terms	**5.813e-04*** (5.542e-05)**	**5.567e-04*** (5.612e-05)**	**5.643e-04*** (5.523e-05)**	**5.456e-04*** (5.585e-05)**
year	**8.795e-04** (3.046e-04)**	6.117e-04 (3.140e-04)	**6.786e-04* (3.108e-04)**	**6.922e-04* (3.122e-04)**
funding	**9.455e-10* (4.294e-10)**	7.427e-10 (4.525E-10)	**9.120e-10* (4.398e-10)**	7.219e-10 (4.510e-10)
trl_level	**4.882e-03* (2.129e-03)**	**4.523e-03* (2.236E-03)**	**5.266e-03* (2.121e-03)**	**4.792e-03* (2.206e-03)**
trl_level_squared	**-1.183e-03*** (3.349e-04)**	**-1.167e-03*** (3.516E-04)**	**-1.245e-03*** (3.340e-04)**	**-1.184e-03*** (3.481e-04)**
tech_proximity		-1.219e-03 (9.862E-04)		
org_balance			1.592e-03 (9.361e-04)	
geo_distance				1.068e-03 (4.584e-04)
geo_distance_sq				**4.747e-03* (1.836e-04)**
Fit	R^2 = .282** 95% CI[.22,.33]	R^2 = .280** 95% CI[.21,.33]	R^2 = .278** 95% CI[.21,.33]	R^2 = .295** 95% CI[.22,.35]
AIC	-3666.894	-3396.541	-3575.606	-3398.575

* indicates $p < .05$. ** indicates $p < .01$. *** indicates p <.001

4.3.2 Structural novelty per IEA category

A possible explanation for the absence of any significant relationships between our explanatory and dependent variables might be that the aggregated model fails to account for inter-group differences in the data. Since different technological fields have their own vocabulary, certain technologies score by default higher on technological novelty than others (see Figure X). Furthermore, the composition of industries differs between technologies. Oil & gas technologies are historically well developed in The Netherlands which could indicate that the oil and gas sector typically consists of incumbents, whereas PV technology is still in its growth phase and therefore allows more new entrants. This could introduce inter-group differences for the proximity related predictors as well. Aggregating these projects into one regression analysis might therefore mask some existing relationships. To rule out any inter-technological differences, our analysis is repeated for individual groups, taking the energy system IEA categories as group levels. We limit our analyses to those IEA categories that were identified by the Dutch government as having the most technological potential for The Netherlands based on expertise present in Dutch knowledge institutes and industries (Janssen et al., 2017). These categories are Solar, Wind, Biofuels and Energy Efficiency projects. We iteratively add variables for individual group level regressions, verifying after each addition whether the predictor leads to an improvement of model quality (AIC). For neither IEA category technological proximity nor geographical proximity improved the fitted models' AIC after being added. For the significant models, a Breusch-Pagan test confirmed that the fitted models were not subject to heteroscedasticity of residuals.

Table 4: Regression results Energy Efficiency Industry using structural novelty as dependent variable

Predictor	**Efficiency Model 0A**	**Efficiency Model 1A**	**Efficiency Model 2A**	**Efficiency Model 3A**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-4.491e-01 (3.423)	1.838e-01 (3.655)	-1.412 (3.091)	1.475e-01 (3.173)
unique_terms	2.133e-04 (1.971e-04)	2.233e-04 (1.999e-04)	1.178e-04 (1.796e-04)	7.134e-05 (1.855e-04)
year	4.428e-04 (1.699e-03)	1.292e-04 (1.814e-03)	9.156e-04 (1.534e-03)	1.451e-04 (1.575e-03)
funding	-1.416e-09 (2.723e-09)	-1.806e-09 (2.845e-09)	-2.632e-09 (2.477e-09)	-1.163e-09 (2.576e-09)
trl_level	-4.872e-03 (5.746e-03)	-5.516e-03 (5.926e-03)	1.091e-05 (5.391e-03)	-4.413e-03 (5.303e-03)
trl_level_squared	1.435e-04 (9.326e-04)	2.419e-04 (9.597e-04)	-5.003e-04 (8.626e-04)	-2.262e-05 (8.569e-04)
tech_proximity		-2.410e-03 (4.525e-03)		
org_balance			**8.491e-03** (2.706e-03)**	
geo_distance				4.171e-03 (2.455e-03)
geo_distance_sq				**-5.646e-04* (2.445e-04)**
Fit	R^2 = .284* 95% CI[.00,.42]	R^2 = .290* 95% CI[.00,.41]	R^2 = .438** 95% CI[.10,.55]	R^2 = .402** 95% CI[.04,.50]
AIC	-293.1196	-291.4572	-301.5135	-292.2573

Table 5: Regression results Wind Energy projects, using structural novelty as dependent variable.

Predictor	**Wind Model 0A**	**Wind Model 1A**	**Wind Model 2A**	**Wind Model 3A**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-1.020 (2.169)	-5.778e-01 (2.200)	1.182 (2.002)	-1.250 (2.286)
unique_terms	**4.099e-04*** **(1.585e-04)**	**4.283e-04*** **(1.590e-04)**	**3.875e-04**** **(1.394e-04)**	**4.225e-04*** **(1.645e-04)**
year	7.187e-04 (1.074e-03)	5.000e-04 (1.089e-03)	-3.729e-04 (9.911e-04)	8.324e-04 (1.131e-03)
funding	-1.069e-09 (2.573e-09)	-1.202e-09 (2.569e-09)	-4.559e-09 (2.461e-09)	-9.562e-10 (2.656e-09)
trl_level	5.783e-03 (2.025e-02)	4.233e-03 (2.025e-02)	-3.220e-03 (1.797e-02)	6.497e-03 (2.091e-02)
trl_level_squared	-2.611e-03 (4.308e-03)	-2.279e-03 (4.308e-03)	1.993e-04 (3.865e-03)	-2.767e-03 (4.456e-03)
tech_proximity		-2.812e-03 (2.565e-03)		
org_balance			**1.020e-02***** **(2.849e-03)**	
geo_distance				9.924e-04 (4.305e-03)
geo_distance_sq				-2.223e-04 (6.499e-04)
Fit	R^2 = .312** 95% CI[.03,.44]	R^2 = .333* 95% CI[.02,.45]	R^2 = .482** 95% CI[.16,.58]	R^2 = .317* 95% CI[.00,.42]
AIC	-314.0078	-313.4039	-325.0904	-310.3252

Results for Energy Efficiency projects are shown in Table 4. We find a positive significant relationship between organizational balance and structural novelty ($p<0.01$), indicating that for Energy Efficiency projects having a balanced set of organization types in the consortium is positively associated with structural novelty. Although we find that the quadratic term of geographical distance is significantly correlated with the dependent variable ($p<0.05$), the same does not hold for the regular term. Hence, we do not find support for the hypothesized inverted u-shaped relationship between geographical proximity and structural novelty in Energy Efficiency projects. Table 5 presents the regression results for Wind Energy projects. Just as for Energy Efficiency projects, we find that organizational balance is significantly correlated to structural novelty ($p<0.001$), which suggests that structural novelty is positively associated with the evenness of organization types in Wind Energy projects. We therefore accept hypothesis 3A for Wind Energy and Energy Efficiency projects and reject hypotheses 1A and 2A

To determine the added value of using an entropy-based measure for organizational proximity as a predictor we fit additional models, using dummies for consortia having an SME or a knowledge institute. We find that for Wind Energy projects having a knowledge institute is significantly related to structural novelty. The model's AIC is slightly higher compared to using organizational balance. This suggests that using entropy based organizational balance does have a slight added value over using a dummy variable. Furthermore, we find that the *number* of knowledge institutes in a consortium is positively related to structural novelty ($p<0.05$). For Energy Efficiency projects we do not find any model improvements after using dummy variables, which indicates that a balanced presence of different organizational arrangements in a consortium is of added value here as well.

For Solar Energy (Table 6) and Biofuels projects (Table 7) neither of the hypothesized predictor variables improved model quality. For both categories, only the number of unique terms in the documents was found to be significantly correlated to structural novelty. This leads us to reject the hypothesized relationships 1A, 2A and 3A for both IEA categories.

Table 6: Regression results Solar Energy projects, using structural novelty as dependent variable.

Predictor	**Solar Model 0A**	**Solar Model 1A**	**Solar Model 2A**	**Solar Model 3A**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-5.033e-02 (1.291)	3.902e-01 (1.363)	1.653e-01 (1.300)	4.336e-02 (1.316)
unique_terms	**6.924e-04***** **(1.237e-04)**	**6.911e-04***** **(1.237e-04)**	**7.036e-04***** **(1.237e-04)**	**6.932e-04***** **(1.251e-04)**
year	2.351e-04 (6.425e-04)	1.626e-05 (6.782e-04)	1.298e-04 (6.465e-04)	1.915e-04 (6.539e-04)
funding	2.485e-09 (1.321e-09)	2.461e-09 (1.321e-09)	2.197e-09 (1.339e-09)	2.384e-09 (1.368e-09)
trl_level	-1.203e-04 (1.137e-02)	-3.025e-04 (1.137e-02)	-2.408e-04 (1.134e-02)	-6.669e-04 (1.175e-02)
trl_level_squared	-4.835e-04 (1.793e-03)	-4.395e-04 (1.793e-03)	-5.070e-04 (1.788e-03)	-4.050e-04 (1.845e-03)
tech_proximity		2.427e-03 (2.411e-03)		
org_balance			-3.844e-03 (3.147e-03)	
geo_distance				-2.153e-03 (4.041e-03)
geo_distance_sq				2.280e-04 (4.042e-04)
Fit	R^2 = .386** 95% CI[.20,.49]	R^2 = .393** 95% CI[.20,.49]	R^2 = .397** 95% CI[.20,.49]	R^2 = .388** 95% CI[.18,.48]
AIC	-644.8722	-643.9602	-644.4706	-641.2212

Table 7: Regression results Biofuels projects, using structural novelty as dependent variable

Predictor	Biofuels Model 0A	Biofuels Model 1A	Biofuels Model 2A	Biofuels Model 3A
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-1.984 (2.173)	-1.960 (2.192)	-2.626 (2.284)	-2.882 (2.285)
unique_terms	**3.705e-04**** **(1.387e-04)**	**3.648e-04*** **(1.422e-04)**	**3.684e-04*** **(1.389e-04)**	**3.788e-04**** **(1.392e-04)**
year	1.189e-03 (1.081e-03)	1.177e-03 (1.091e-03)	1.509e-03 (1.136e-03)	1.638e-03 (1.137e-03)
funding	-2.559e-10 (6.843e-10)	-2.779e-10 (6.969e-10)	-3.499e-10 (6.927e-10)	-2.496e-10 (6.886e-10)
trl_level	1.398e-02 (1.653e-02)	1.453e-02 (1.684e-02)	1.144e-02 (1.678e-02)	1.117e-02 (1.672e-02)
trl_level_squared	-2.956e-03 (3.449e-03)	-3.049e-03 (3.501e-03)	-2.394e-03 (3.506e-03)	-2.390e-03 (3.492e-03)
tech_proximity		5.250e-04 (2.383e-03)		
org_balance			1.978e-03 (2.142e-03)	
geo_distance				-1.482e-03 (3.442e-03)
geo_distance_sq				2.191e-04 (3.174e-04)
Fit	R^2 = .164* 95% CI[.00,.27]	R^2 = .165 95% CI[.00,.26]	R^2 = .176 95% CI[.00,.28]	R^2 = .188 95% CI[.00,.28]
AIC	-466.8042	-464.8582	-465.7482	-464.738

4.4 Regression results dependent variable II: functional novelty

This part of the analysis focuses on the second dependent variable, functional novelty. Functional novelty pertains to the degree to which projects are different compared to preceding projects.

4.4.1 Aggregated model

The results for the aggregated model, in which a project's abstract is compared to all preceding projects, are presented in Table 8. We find that adding organizational distance as a predictor variable improves the model quality, whereas the other proximity dimensions do not. Therefore, we confirm hypothesis 3B and reject hypotheses 1B and 2B. Just as with structural novelty, the number of unique words in an abstract is significantly correlated with functional novelty. The opposite is true for the project’s starting year. This is intuitive since functional novelty of a document is measured in comparison to what documents existed before. When a project’s starting year is higher, it is measured against more preceding documents, limiting the probability that it adds novelty.

Table 8: Regression results using functional novelty as dependent variable

Predictor	Model 0B	Model 1B	Model 2B	Model 3B
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	**1.145*** (2.384e-01)**	**- 1.410*** (2.460e-01)**	**1.155*** (2.384e-01)**	**1.182*** (2.376e-01)**
unique_terms	**2.879e-04*** (2.183e-05)**	**3.005e-04*** (2.190e-05)**	**3.017e-04*** (2.186e-05)**	**3.040e-04*** (2.173e-05)**
year	**-3.487-04** (1.185e-04)**	**-4.805e-04*** (1.222e-04)**	**-3.543e-04** (1.185e-04)***	**-3.678e-04** (1.181e-04)**
funding	1.674e-10 (1.831e-10)	9.188e-11 (1.797-10)	1.348e-10 (1.847e-10)	1.381e-10 (1.824e-10)
trl_level	**3.260e-03*** (8.598e-03)**	**3.370e-03 (8.755e-04)***	**3.418e-03*** (8.585e-04)**	**3.409e-03*** (8.446e-04)**
trl_level_squared	**-5.902e-04*** (1.367e-04)**	**-6.256e-04 (1.378e-04)***	**6.053e-04*** (1.368e-04)**	**-6.175e-04*** (1.344e-04)**
tech_proximity		-1.128e-03 (6.154e-04)		
org_balance			**7.613e-04* (3.715e-04)**	
geo_distance				2.750e-04 (1.476e-04)
Fit	R^2 = .314** 95% CI[.24,.37]	R^2 = .347** 95% CI[.27,.40]	R^2 = .327** 95% CI[.25,.38]	R^2 = .351** 95% CI[.28,.40]
AIC	-4191.975	-3991.085	-4209.359	-4164.073

From the fitted models we find that organizational balance is significantly correlated with functional novelty ($p<0.05$), indicating that TSE projects carried out by a balanced set of organization types are associated with higher functional novelty. We further inspect this relationship by fitting (dummy) variants of organizational proximity, representing different aspects of organizational proximity. This allows us to gain insights into which specific factors are contributing to the observed relationship. We test the presence of an SME in the consortium, the presence of a knowledge institute and the number of knowledge institutes. We find that both including a knowledge institute and having more knowledge institutes significantly contributes to functional novelty ($p<0.05$). Including an SME was not found to significantly contribute to functional novelty. To further confirm the findings related to organizational proximity and functional novelty, we repeated the aggregated analysis using expert's ratings of innovativeness as the dependent variable. Although expert ratings were only available for half of the projects, the analysis confirmed that there is a significant relationship between organizational balance and innovativeness ($p<0.001$), after controlling for year, funding, and TRL level. The results can be found in Appendix C. The other proximity dimensions were not found to be significantly related to expert innovativeness rating. Lastly, we performed a simpler operationalization of organizational balance by counting the number of unique organization types in a project. This confirmed the results and led to an even better model AIC.

4.4.2 Functional novelty per IEA category

Similar to structural novelty, we fit linear regression models for functional novelty per IEA category to uncover potentially masked relationships. To this end, we recompute functional novelty taking into account only the preceding projects within that IEA category to compare newly introduced projects against. We discard any IEA categories smaller than 50 projects and remove outliers based on Cook's distance.

Table 9: Regression results Energy Efficiency Industry, using functional novelty as the dependent variable

Predictor	**Efficiency Model 0B**	**Efficiency Model 1B**	**Efficiency Model 2B**	**Efficiency Model 3B**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-2.071 (2.074)	-2.116 (2.151)	-2.205 (1.890)	-1.830 (2.139)
unique_terms	1.534e-04 (1.115e-04)	1.523e-04 (1.136e-04)	**2.747e-04*** **(1.160e-04)**	1.812-04 (1.148e-04)
year	1.251e-03 (1.030e-03)	1.273e-03 (1.068e-03)	1.319e-03 (9.382e-04)	1.131e-03 (1.062e-03)
funding	**-3.563e-09*** **(1.485e-09)**	**-3.548e-09*** **(1.513e-09)**	-2.706e-09 (1.416e-09)	**-3.927e-09*** **(1.528e-09)**
trl_level	**-7.330e-03*** **(3.421e-03)**	-7.244e-03 (3.575e-03)	**-9.290e-03**** **(3.246e-03)**	**-8.022e-03*** **(3.498e-03)**
trl_level_squared	**1.547e-03*** **(5.709e-04)**	**1.529e-03*** **(6.058e-04)**	**1.757e-03**** **(5.340e-04)**	**1.670e-03*** **(5.856e-04)**
tech_proximity		2.083e-04 (2.069e-03)		
org_balance			**-5.702e-03**** **(1.770e-03)**	
geo_distance				1.248e-04 (6.079e-04)
Fit	R^2 = .367** 95% CI[.05,.49]	R^2 = .367** 95% CI[.03,.48]	R^2 = .538** 95% CI[.21,.63]	R^2 = .393** 95% CI[.05,.51]
AIC	-331.3266	-329.3388	-336.9222	-321.8289

The results of the regression analysis for Energy Efficiency projects are presented in Table 9. The high R-squared value of the model suggests that a significant portion of the variance in functional novelty can be explained by the predictor variables. Interestingly, there is a significant *negative* correlation between organizational balance and functional novelty ($p<0.01$), indicating that projects from consortiums that are organizationally distant tend to have lower levels of functional novelty. This opposes our findings for *structural* novelty in Energy Efficiency projects (Table 4) and leads us to reject hypotheses 1B, 2B and 3B. Additionally, the relationship between TRL level and functional novelty is found to have a U-shaped pattern ($p<0.01$), which opposes the *inverted* U-shaped relationship observed in the aggregated model. Further analysis of the documents within the group reveals that projects involving Biorefinery, (molecular) Separation, and Chemical Catalytic Conversion contribute to low functional novelty scores at TRL 2 (Development).

Table 10: Regression results Solar Energy, using functional novelty as the dependent variable

Predictor	**Solar Model 0B** *b (Std. Error)*	**Solar Model 1B** *b (Std. Error)*	**Solar Model 2B** *b (Std. Error)*	**Solar Model 3B** *b (Std. Error)*	**Solar Model 4B** *B (Std. Error)*
(Intercept)	1.697e-01 (1.259)	-7.855e-01 (1.275e+00)	-9.700e-02 (1.238)	1.646e-01 (1.265)	-9.17e-01 (1.26)
unique_terms	-3.482e-05 (1.183e-04)	-3.184e-05 (1.146e-04)	-5.032e-05 (1.160e-04)	-3.095e-05 (1.194e-04)	-4.54e-05 (1.13e-04)
year	1.279e-04 (6.256e-04)	6.042e-04 (6.339e-04)	2.590e-04 (6.152e-04)	1.293e-04 (6.289e-04)	6.68e-04 (6.25e-04)
funding	-9.067e-10 (1.021e-09)	-8.652e-10 (9.893e-10)	-1.274e-09 (1.014e-09)	-9.255e-10 (1.028e-09)	-1.18e-09 (9.89e-10)
trl_level	8.050e-03 (1.543e-02)	5.652e-03 (1.497e-02)	6.308e-03 (1.512e-02)	8.248e-03 (1.552e-02)	4.40e-03 (1.48e-02)
trl_level_squared	-9.883e-04 (2.526e-03)	-6.018e-04 (2.451e-03)	-6.733e-04 (2.476e-03)	-1.004e-03 (2.540e-03)	-3.71e-04 (2.42e-03)
tech_proximity		**-5.627e-03* (2.198e-03)**			**-5.06e-03* (2.19e-03)**
org_balance			**6.526e-03* (3.001e-03)**		5.59e-03 (2.95e-03)
geo_distance				3.760e-04 (1.026e-03)	
Fit	R^2 = .047 95% CI[.00,.10]	R^2 = .117 95% CI[.00,.20]	R^2 = .099 95% CI[.00,.17]	R^2 = .049 95% CI[.00,.09]	R^2 = .154* 95% CI[.00,.23]
AIC	-630.06	-634.8988	-633.0483	-628.2056	-636.7416

For Solar Energy projects (Table 10), we find a significant negative relationship between technological proximity and functional novelty. This suggests that Solar Energy projects contribute more to functional novelty if project partners are technologically diverse. We therefore accept Hypothesis 1B for Solar Energy projects. We test whether this relationship also exists when using a more basic operationalization of technological proximity, i.e., the number of unique SIC codes in a project consortium. This relationship is not significant, which could indicate that using a continuous measure of differences between SIC codes captures additional information of technological relatedness between organizations compared to the unique SIC count (Stirling, 2007). We further find that organizational balance is significantly correlated in model 2B but loses significance (p=0.06) when we control for technological proximity (model 4B). Potentially, the predictors technological proximity and organizational balance show slight multicollinearity for Solar Energy projects. However, the associated VIF value did not suggest significant multicollinearity. Due to the superior model fit of model 1B compared to model 2B, we reject hypothesis 2B for Solar Energy projects. Geographical distance is not found to be significant (model 3B), which leads us to reject hypothesis 3B for solar energy projects as well.

Table 11: Regression results Wind Energy, using functional novelty as the dependent variable

Predictor	**Wind Model 0B**	**Wind model 1B**	**Wind Model 2B**	**Wind Model 3B**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	-1.084e-01 (2.003)	-8.087e-01 (1.957)	-5.860e-01 (2.073)	3.012e-01 (1.879)
unique_terms	-9.012e-05 (1.368e-04)	-9.273e-05 (1.316e-04)	-7.936e-05 (1.376e-04)	-1.017e-04 (1.279-04)
year	2.642e-04 (1.001e-03)	6.12e-04 (9.78e-04)	5.047e-04 (1.036e-03)	5.671e-05 (9.389e-04)
funding	**-8.608e-09** (2.507e-09)**	**-7.55e-09** (2.47e-09)**	**-7.563e-09** (2.756e-09)**	**-8.529e-09*** (2.343e-09)**
trl_level	7.249e-03 (2.418e-02)	6.60e-03 (2.33e-02)	3.074e-03 (2.465e-02)	6.804e-03 (2.260e-02)
trl_level_squared	-1.104e-03 (4.455e-03)	-9.94e-04 (4.29e-03)	-5.017e-04 (4.512e-03)	-1.096e-03 (4.164e-03)
tech_proximity		4.38e-03 (2.17e-03)		
org_balance			-2.808e-03 (3.049e-03)	
geo_distance				**2.329e-03* (9.133e-04)**
Fit	*R2* = .294* 95% CI[.01,.42]	*R2* = .364** 95% CI[.04,.48]	*R2* = .310* 95% CI[.00,.43]	*R2* = .399** 95% CI[.07,.51]
AIC	-311.0587	-313.6669	-310.0557	-316.1829

For Wind Energy projects (Table 11), we find geographical distance to be significantly correlated to functional novelty ($p<0.05$, model 3B). This suggests that higher geographical distance between R&D project partners contributes to higher functional novelty of wind energy projects. We therefore accept hypothesis 3B for Wind Energy projects. We further investigate this relationship by testing whether using a dummy variable indicating a foreign participant in the R&D project preserves the relationship's significance. The significance was lost when repeating the analysis with a dummy. Other predictors were not found to be significant, meaning that we reject hypotheses 1B and 2B for wind energy projects. Interestingly, funding is strongly negatively associated with functional novelty. A potential explanation for this could be that projects exploring technologically safe directions with a higher probability of success are more likely to secure substantial funding.

Table 12: Regression results Biofuels, using functional novelty as the dependent variable

Predictor	**Biofuels Model 0B**	**Biofuels Model 1B**	**Biofuels Model 2B**	**Biofuels Model 3B**
	b (Std. Error)	*b (Std. Error)*	*b (Std. Error)*	*b (Std. Error)*
(Intercept)	5.666e-01 (1.486)	5.527e-01 (1.494)	2.373e-01 (1.534)	4.086e-01 (1.580)
unique_terms	**3.207e-04**** **(9.619^{e}-05)**	**3.103e-04**** **(9.815^{e}-05)**	**3.245e-04**** **(9.645^{e}-05)**	**3.230e-04**** **(9.721^{e}-05)**
year	-6.111e-05 (7.388^{e}-04)	-5.439e-05 (7.428^{e}-04)	1.025e-04 (7.624^{e}-04)	1.721e-05 (7.851^{e}-04)
funding	6.522e-10 (4.322^{e}-10)	6.056e-10 (4.410^{e}-10)	5.888e-10 (4.388^{e}-10)	6.509e-10 (4.356^{e}-10)
trl_level	-8.682e-03 (7.779^{e}-03)	-8.392e-03 (7.834^{e}-03)	-9.371e-03 (7.830^{e}-03)	-9.150e-03 (7.979^{e}-03)
trl_level_squared	1.325e-03 (1.564^{e}-03)	1.304e-03 (1.572^{e}-03)	1.473e-03 (1.575^{e}-03)	1.428e-03 (1.609^{e}-03)
tech_proximity		9.442e-04 (1.531^{e}-03)		
org_balance			1.242e-03 (1.391^{e}-03)	
geo_distance				1.587e-04 (5.052e-04)
Fit	R^2 = .296** 95% CI[.07,.41]	R^2 = .300** 95% CI[.06,.41]	R^2 = .305** 95% CI[.06,.41]	R^2 = .297** 95% CI[.06,.41]
AIC	-504.7388	-503.1634	-503.6256	-502.8493

For biofuels projects (Table 12), the addition of any hypothesized predictor variable leads to higher AIC values, indicating a decrease in model fit. We therefore reject hypotheses 1B, 2B and 3B for biofuels projects.

5. Discussion

This section outlines the findings as well as academic and practical implications of the conducted analyses. It concludes with the limitations of the methodology and suggestions for further research.

5.1 Academic implications

Structural novelty

For Wind energy and Energy Efficiency technologies our results support the idea that the creation of structural novelty is a collaborative act that benefits from organizational balance in R&D projects. Thereby, we complement existing literature on university-industry-government relations, such as the National Innovation Systems (Lundvall, 1992) and Triple Helix thesis (Etzkowitz & Leydesdorff, 2000). Beyond the known societal benefits of university-industry linkages such as efficient transfer of technology and knowledge spillovers, our results indicate that university-industry linkages also contribute to increased *technological novelty* compared to inter-industry linkages. Thereby, our results seem to point in the same direction as Talab et al. (2020), who show that during interorganizational collaboration universities are most active in the development of new ideas. In the same vein, Tanner (2014) finds in her study on fuel cell technology that technological diversification heavily relies on knowledge from nonindustrial actors such as knowledge institutes and universities.

The difference in accepted hypotheses between IEA categories might imply there is a fundamental difference between IEA subgroups. The question is whether this can be attributed to the inherent properties of those IEA subgroups or to an underlying confounding variable that is present for two of the groups and absent for the two other groups. An ad hoc explanation of these differences can be made based on Klepper's Industry Life Cycle (Klepper, 1997) that describes the different stages of industry development. As industries mature, the focus of innovative activities typically shifts from radical product innovation towards refining existing processes and improving efficiency. However, empirical studies have demonstrated that for many technologies there is no decline in product innovations over time (Gort & Klepper, 1982; Henderson et al., 1995; Lee & Berente, 2013). Between solar, biofuel and wind technology Huenteler et al. (2016) identify a strong difference regarding the shift in innovative activities. Specifically, they observe that in the wind energy sector the focus of innovation has shifted across various components of the product, rather than a transition towards process innovation. Biofuel and Solar technology on the other hand are characterized by a higher scale of production process and lower complexity of product architecture (Huenteler et al., 2016). These industries follow the more typical pattern where focus shifts from early product innovations to process innovations for solar cell mass production. The former cycle might also apply to Energy Efficiency technologies as this cycle is more credible for technologies with multiple potential applications and markets (Winskel et al., 2014). Given the difference in the nature of innovation, the role of organizational proximity in the creation of technological novelty may vary across different IEA groups. Another potential explanation could be the difference in knowledge bases between IEA subgroups. In the nano-biopharmaceutical field Zhang & Tang (2018) find that the effect of organizational balance on innovation performance is moderated by structural holes of knowledge elements from R&D partners and the degree centrality of an organization's knowledge elements. Energy Efficiency and Wind Energy projects might exhibit a higher degree of structural holes and degree centrality of knowledge elements compared to solar and biofuels projects. Future research may focus on studying the impact of both industry life cycle types and structural holes on this relationship.

We find no evidence for the effect of geographical proximity on structural novelty, which might seem surprising as it is one of the most widely studied concepts of the proximities framework. However, Boschma (2005) argues that geographical proximity is not a sufficient nor a necessary condition for knowledge creation to take place. It may create favorable conditions but other factors, such as shared interests, trust, social networks, and institutional support are more important. Furthermore, interactive learning is not the dependent variable in this study, which might explain the absence of any significant relationships. Another explanation for the absence of the relationship might be the limited size of The Netherlands or the limited participation of non-Dutch organizations in R&D projects. 95% of the TSE projects were executed by Dutch organizations only (Appendix B). For solar technology, Li et al (2021) find that unrelated technologies are more likely combined when they originate from the same region compared to internationally. Potentially our data contained an insufficient number of internationally oriented projects to uncover such a relationship.

Functional novelty

We find evidence for a relationship between organizational proximity and functional novelty. These results are in line with findings by Nieto & Santamaria (2007), who find that collaborative networks comprising different partner types have a positive effect on functional novelty. Regarding the presence of research organizations in the consortium our results seem to confirm previous research that industry-university collaboration in R&D projects yields more novel outcomes (Cassiman et al., 2010). Our aggregated functional novelty measure is similar to the technological diversity measure proposed by Paez-Aviles et al. (2018), who also study its relationship to project team variables. Interestingly, our results deviate from theirs. We find that the balance of project partners in R&D collaboration is associated with higher functional novelty (Table 8), whereas they do not find support for this. On the other hand, they find strong evidence for the role of technological and geographical proximity, whereas we only find support for that in some IEA categories. A potential explanation could be the difference in focus of the two studies. Our focus is on energy technology whereas their focus is on nanotechnology projects. Furthermore, the discrepancy of results is possibly attributable to differences in operationalization. We compute a continuous novelty measure based on an abstract's similarity to previously existing abstracts, whereas Paez Aviles et al. (2018) compute added technological diversity using a categorical based entropy measure. For knowledge base they take the number of patents whereas we again use a similarity measure based on relatedness of SIC codes. Further research could apply our methodology to their data or vice versa to verify whether operationalization differences can explain the deviations in results.

Just as with structural novelty, our results show that the hypothesized relationships differ per IEA category for functional novelty as well. For Solar Energy, technological proximity is negatively associated with functional novelty. This is consistent with theory arguing that radical innovation likely emerges from the combination of unrelated knowledge bases (Janssen & Frenken, 2019). On the other hand, too much technological proximity between knowledge bases leads to more incremental technological novelty (Miguelez & Moreno, 2018). For wind energy projects geographical distance was found to be positively associated with functional novelty. This seems to indicate that larger distances between project partners help to develop technologies that deviate from existing practices. Since foreign participation does not appear to contribute to this relationship, regional specialization within The Netherlands potentially enhances the functional novelty of wind energy projects.

For Energy Efficiency Industry (EEI) projects, organizational proximity (balance) is positively (negatively) associated with functional novelty. The contradictory relationships between structural and functional novelty of organizational proximity in Energy Efficiency Industry projects might be attributed to the difference in nature between structural and functional novelty. Potentially, too much organizational balance poses challenges for generating functional novelty in EEI projects whereas it fosters structural novelty. As functional novelty is more oriented towards societal needs and market application (Argyres and Silverman, 2004; Arthur, 2007) too much organizational balance between project partners might hinder the convergence of their views towards a technological design that effectively addresses a societal need. An alternative explanation could be the technology-transcending nature of EEI projects, which can be observed from the dispersion of green data points in Figure 8. As EEI technologies can encompass multiple functionalities of the energy system, this IEA category might not lend itself well for a cross-comparison of its projects. This difference between IEA categories can also be observed in Figure 11. The two measures of technological novelty are positively correlated (p<0.001) for non-EEI projects, but this relationship is not present for EEI projects. Indeed, previous studies on functional novelty and its relationship with team characteristics take a more technology-narrow approach (e.g. Van Rijnsoever et al., 2015; Paez-Avlies et al., 2018 for biogas and nanotechnology respectively).

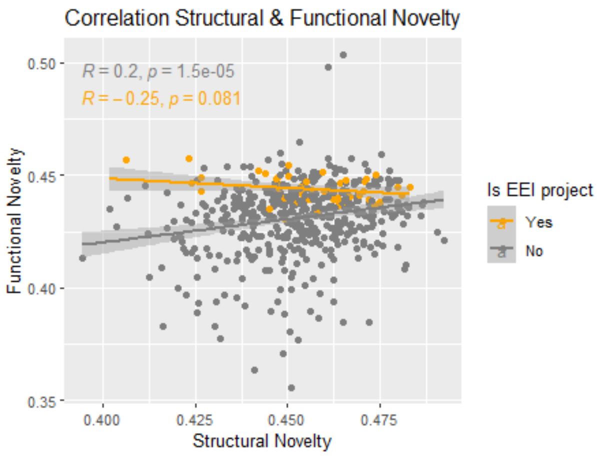


Figure 11: Correlation structural and functional novelty: EEI projects (orange) vs other projects.

5.2 Practical Implications

Our results have implications for policy makers and practitioners in The Netherlands such as TKI managers[6]. Primarily, the hypothesized relationships between proximities and technological novelty were found to differ across energy technologies. Our findings therefore emphasize the necessity of tailoring energy innovation policy to individual technologies through bodies such as TKIs to account for intersectoral differences, rather than pursuing generic energy innovation policy.

Regarding the role of organizational proximity, we find some evidence that supports current Dutch energy policy in its aim to bring together knowledge institutes and industry to foster

6 TKI managers are in charge of agenda setting on behalf of government-industry-academia in The Netherlands.

energy innovation. Policy makers could consider involving multiple knowledge institutes in R&D projects to further enhance the generation of structural novelty in Wind and Energy Efficiency projects and functional novelty overall. For Energy Efficiency projects, however, this relationship was the opposite. For Solar Energy specifically, our results could assist TKI managers to coordinate the search for project participants. It may help to consider the technological distance between potential R&D partners if the aim is to increase a project's functional novelty. In the same vein, one could consider the geographical distance among R&D project participants to enhance the creation of technological diversity in the field of wind energy,

One should note however that while reducing participant proximities can foster technological novelty, some scholars argue that too much novelty has its disadvantages. On the neoclassical side of the academic arena, too much technological diversity is believed to impede (economic) efficiency. Higher diversity purportedly leads to increased production costs, lower economies of scale and hampers product standardization (Ricardo, 1817; Cohendet & Llerena, 1992). Furthermore, engaging in radically different activities introduces new routines that require learning which might cause coordination inefficiencies between economic actors (Cohendet & Llerena, 1997). These disadvantages of collaborative R&D are attributable to increased transaction costs (Williamson, 1987).

Secondly, the text-based approach to measure technological novelty provided intuitive results. Our methodology could therefore serve as a tool for TKI managers or innovation policy consultants to quantify loosely defined government objectives such as "more innovation" (Janssen et al., 2017, p5). It allows for objectively assessing the relative novelty of R&D projects and could supplement expert ratings when determining eligibility for funding. In the same vein, the approach can be applied to evaluate R&D policy and assess the R&D projects' contribution to technological novelty in the innovation system. Preferably, this would require accurate and standardized documentation of R&D project proposals and results.

5.3 Limitations and further research

This study has several limitations that are important to take into account when interpreting the results. First of all, regressing the proposed operationalizations of technological novelty against known measures of novelty provided only limited evidence for their validity. Regarding validation we find that the energy subgroups associated with high and low technological novelty are intuitive. Furthermore, we found that *structural* novelty is significantly correlated with the expert assessment of a project's technological continuity; a binary variable indicating whether a project deviates from prevailing technological practices. For *functional* novelty however, we did not find such a relationship. A potential methodological explanation for this is that standardization of project descriptions changes over time, which results in larger differences between documents than desirable. For instance, some abstracts report on the consortium whereas others emphasize technological challenges. This creates undesirable discrepancies between descriptions and hampers obtaining the project's true technological novelty. Further research could account for these discrepancies by using more standardized documents, such as patents. An interesting direction could involve regressing the novelty of patent claims against proximity. Alternatively, future research could aim to train a custom sentence or paragraph classifier that is able to distinguish technology related sections from non-technology related ones in R&D project descriptions.

Secondly, word vectors have a time dependency which influences the accuracy of our novelty measure. Technological artifacts become more or less related over time as a result of technological development. For instance, batteries and automotive technology have become increasingly related over the past 20 years. Although we control for time in the regression models, using point-in-time word vectors would provide a more accurate representation of technological novelty. This will ignore any knowledge that was published after the time of publication of the document that technological novelty is computed over. To solve this issue, we experimented with training custom word vectors for every project starting year in our dataset, based on a set of 2.5 million US Energy patents. The number of patents however proved insufficient to cover enough vocabulary of the TSE project abstracts to be able to accurately compute technological novelty. Furthermore, the performance of our custom trained word vectors on validation tasks (analogy test and similarity test) was far below that of the pretrained GloVe model (see Appendix D). Future research could obtain accurate point-in-time word vectors by training over a much larger set of documents, covering sufficient vocabulary.

6. Conclusion

Despite widespread consensus among scholars regarding the importance of technological novelty in driving economic development and addressing societal challenges, still very little is known about the drivers that underly its creation. Other studies in the field of evolutionary economics qualitatively explain case-specific trajectories of diversity in technological designs but do not empirically assess the drivers that have generated it (Cohendet & Llerena, 1997; Frenken & Leydesdorff, 2000). The present study quantitively examined the drivers of technological novelty in collaborative R&D projects. We developed a text-based approach to quantify technological novelty of energy R&D projects. We then applied this measure to analyse which proximity dimensions of project consortia predict technological novelty creation. In particular, we studied in what way technological, geographical and organizational proximity are associated with structural and functional technological novelty, through the following research question:

How does proximity between project partners relate to the creation of technological novelty in collaborative R&D projects?

Regarding the role of proximity dimensions on the creation of technological novelty we find mixed and in the case of Energy Efficiency projects seemingly contrasting results. We find some evidence that technological proximity and organizational proximity influence technological novelty, each in a separate way. For Wind and Energy Efficiency projects, organizational proximity was found to contribute to structural novelty. For Energy Efficiency projects, we found organizational proximity to have the opposite effect on functional novelty. For solar and wind energy projects we find that technological proximity and geographical proximity respectively are negatively associated with functional novelty.

The research question can therefore not be answered unequivocally, also possibly due to the explorative nature of our research methodology. We illustrated the potential for using text-based analysis to detect technological novelty of innovation projects. When looking at projects with high and low novelty scores we find the results to be intuitive. Furthermore, the proposed structural novelty measure can be used to distinguish between continuous and discontinuous energy projects. We do however find that the proposed measures are strongly correlated with the number of unique words used in a document and that document standardization is critical. It would be interesting to keep developing the methodological approach using the latest advancements in NLP, thereby improve the quantification of technological novelty.

References

- Aghion, P., & Jaravel, X. (2015). Knowledge spillovers, innovation and growth. *The Economic Journal*, *125*(583), 533-573.
- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, *45*(3), 425-455.
- Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic management journal*, *22*(6-7), 521-543.
- Almeida, P., & Kogut, B. (1997). The exploration of technological diversity and geographic localization in innovation: Start-up firms in the semiconductor industry. *Small Business Economics*, *9*, 21-31.
- Anderson, P., & Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative science quarterly*, 604-633.
- Argyres, N. S., & Silverman, B. S. (2004). R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management Journal*, *25*(8-9), 929-958.
- Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies*, *29*(3), 155-173.
- Arthur, W. B. (2007). The structure of invention. *Research policy*, *36*(2), 274-287.
- Aschhoff, B. (2009). The effect of subsidies on R&D investment and success–Do subsidy history and size matter?. *ZEW-Centre for European Economic Research Discussion Paper*, (09-032).
- Basalla, G. (1988). *The evolution of technology*. Cambridge University Press.
- Baum, J. A., Calabrese, T., & Silverman, B. S. (2000). Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. *Strategic management journal*, *21*(3), 267-294.
- Belderbos, R., Carree, M., & Lokshin, B. (2004a). Cooperative R&D and firm performance. *Research policy*, *33*(10), 1477-1492.
- Belderbos, R., Carree, M., Diederen, B., Lokshin, B., & Veugelers, R. (2004b). Heterogeneity in R&D cooperation strategies. *International journal of industrial organization*, *22*(8-9), 1237-1263.
- Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, *39*(1), 61-74.
- Boschma, R. A., & Lambooy, J. G. (1999). Evolutionary economics and economic geography. *Journal of evolutionary economics*, *9*(4), 411-429.
- Boschma, R., Eriksson, R. H., & Lindgren, U. (2014). Labour market externalities and regional growth in Sweden: The importance of labour mobility between skill-related industries. *Regional Studies*, *48*(10), 1669-1690.
- Branstetter, L. G., & Sakakibara, M. (2002). When do research consortia work well and why? Evidence from Japanese panel data. *American Economic Review*, *92*(1), 143-159.
- Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of economic geography*, *12*(2), 409-433.
- Bunduchi, R. (2013). Trust, partner selection and innovation outcome in collaborative new product development. *Production planning & control*, *24*(2-3), 145-157.
- Capaldo, A., & Messeni Petruzzelli, A. (2015). Origins of knowledge and innovation in R&D alliances: a contingency approach. *Technology Analysis & Strategic Management*, *27*(4), 461-483.
- Capaldo, A., & Petruzzelli, A. M. (2014). Partner geographic and organizational proximity and the innovative performance of knowledge-creating alliances. *European Management Review*, *11*(1), 63-84.
- Carnabuci, G., & Operti, E. (2013). Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic management journal*, *34*(13), 1591-1613.
- Cassiman, B., Di Guardo, M. C., & Valentini, G. (2010). Organizing links with science: Cooperate or contract?: A project-level analysis. *Research Policy*, *39*(7), 882-892.
- Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional studies*, *49*(5), 767-781.
- Chandy, R. K., & Tellis, G. J. (1998). Organizing for radical product innovation: The overlooked role of willingness to cannibalize. *Journal of marketing research*, *35*(4), 474-487.
- Chandy, R. K., & Tellis, G. J. (2000). The incumbent's curse? Incumbency, size, and radical product innovation. *Journal of marketing*, *64*(3), 1-17.
- Christensen, C. (1997). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press. Cambridge, MA, 288p.
- Christensen, C. M. (2013). *The innovator's dilemma: when new technologies cause great firms to fail.* Harvard Business Review Press.
- Cockburn, I. M., & Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial economics*, *46*(2), 157-182.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, *99*(397), 569-596.

- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.
- Cohendet, P., & Llerena, P. (1997). Learning, technical change and public policy: how to create and exploit diversity. *Edquist, C., Systems of innovation, Pinter, London and Washington*, 223-241.
- Cohendet, P., Llerena, P., & Sorge, A. (1992). Technological diversity and coherence in Europe: an analytical overview. *Revue d'économie Industrielle, 59*(1), 9-26.
- Colombo, M. G. (2003). Alliance form: A test of the contractual and competence perspectives. *Strategic management journal*, *24*(12), 1209-1229.
- Cook, R. D., & Weisberg, S. (1982). Criticism and influence analysis in regression. *Sociological methodology*, *13*, 313-361.
- Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis issues for field studies Boston. *MA: Houghton Mifflin*.
- Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and corporate change*, *10*(4), 945-974.
- Cooke, P. (2008). Regional innovation systems, clean technology & Jacobian cluster-platform policies. *Regional Science Policy & Practice*, *1*(1), 23-45.
- Cooke, P., Uranga, M. G., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organizational dimensions. *Research policy*, *26*(4-5), 475-491.
- Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, *11*(3), 147-162.
- Dosi, G. (1988). Sources, procedures, and microeconomic effects of innovation. *Journal of economic literature*, 1120-1171.
- Du, J., Leten, B., & Vanhaverbeke, W. (2014). Managing open innovation projects with science-based and market-based partners. *Research Policy*, *43*(5), 828-840.
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research policy*, *29*(2), 109-123.
- Fabrizio, K. R. (2009). Absorptive capacity and the search for innovation. *Research policy*, *38*(2), 255-267.
- Fitjar, R. D., & Rodríguez-Pose, A. (2014). When local interaction does not suffice: sources of firm innovation in urban Norway. In *Regional development and proximity relations* (pp. 195-222). Edward Elgar Publishing.
- Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117-132.
- Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic management journal*, *25*(8-9), 909-928.
- Florida, R. (2005). The flight of the creative class: The new global competition for talent. *Liberal education*, *92*(3), 22-29.
- Frenken, K., & Leydesdorff, L. (2000). Scaling trajectories in civil aircraft (1913–1997). *Research Policy*, *29*(3), 331-348.
- Galunic, D. C., & Rodan, S. (1998). Resource recombinations in the firm: Knowledge structures and the potential for Schumpeterian innovation. *Strategic management journal*, *19*(12), 1193-1201.
- Gertler, M. S. (1995). “Being there”: proximity, organization, and culture in the development and adoption of advanced manufacturing technologies. *Economic geography*, *71*(1), 1-26.
- Gort, M., & Klepper, S. (1982). Time paths in the diffusion of product innovations. *The economic journal*, *92*(367), 630-653.
- Green, S. G., Gavin, M. B., & Aiman-Smith, L. (1995). Assessing a multidimensional measure of radical technological innovation. *IEEE transactions on engineering management*, *42*(3), 203-214.
- Haghighi Talab, A., Scholten, V., & van Beers, C. (2020). The role of universities in inter-organizational knowledge collaborations. *Journal of the Knowledge Economy*, *11*, 458-478.
- Hamel, G. (1991). Competition for competence and interpartner learning within international strategic alliances. *Strategic management journal*, *12*(S1), 83-103.
- Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American sociological review*, 149-164.
- Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of management review*, *32*(4), 1199-1228.
- Henderson, R. (1995). Of life cycles real and imaginary: The unexpectedly long old age of optical lithography. *Research Policy*, *24*(4), 631-643.
- Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.
- Hill, C. W., & Rothaermel, F. T. (2003). The performance of incumbent firms in the face of radical technological innovation. *Academy of management review*, *28*(2), 257-274.
- Huenteler, J., Schmidt, T. S., Ossenbrink, J., & Hoffmann, V. H. (2016). Technology life-cycles in the energy sector—Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, *104*, 102-121.
- Jacobsson, S., & Bergek, A. (2004). Transforming the energy sector: the evolution of technological systems in renewable energy technology. *Industrial and corporate change*, *13*(5), 815-849.
- Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value.

- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, *108*(3), 577-598.
- Jang, H., Lee, K., & Yoon, B. (2017). Development of an open innovation model for R&D collaboration between large firms and small-medium enterprises (SMES) in manufacturing industries. *International Journal of Innovation Management*, *21*(01), 1750002.
- Janssen, M., de Boer, P. J., Jelicic, N., Hof, B., van der Werff, S., & van den Berg, E. (2017). Beleidsevaluatie Energie-innovatieregelingen. Source: https://www.dialogic.nl/wp-content/uploads/2018/07/Eindrapport-Evaluatie-Energie-innovatieregelingen.pdf
- Janssen, M. J., & Frenken, K. (2019). Cross-specialisation policy: Rationales and options for linking unrelated industries. *Cambridge Journal of Regions, Economy and Society*, *12*(2), 195-212.
- Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of management journal*, *45*(6), 1183-1194.
- Kemp, R. (1994). Technology and the transition to environmental sustainability: the problem of technological regime shifts. *Futures*, *26*(10), 1023-1046.
- Kirat, T., & Lung, Y. (1999). Innovation and proximity: territories as loci of collective learning processes. *European urban and regional studies*, *6*(1), 27-38.
- Klepper, S. (1997). Industry life cycles. *Industrial and corporate change*, *6*(1), 145-182.
- Knoben, J., & Oerlemans, L. A. (2006). Proximity and inter-organizational collaboration: A literature review. *international Journal of management reviews*, *8*(2), 71-89.
- Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.
- Lee, J., & Berente, N. (2013). The era of incremental change in the technology innovation life cycle: An analysis of the automotive emission control industry. *Research Policy*, *42*(8), 1469-1481.
- Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic management journal*, *14*(S2), 95-112.
- Li, D., Heimeriks, G., & Alkemade, F. (2021). Recombinant invention in solar photovoltaic technology: can geographical proximity bridge technological distance?. *Regional Studies*, *55*(4), 605-616.
- Lundvall, B.A. (1992). *National Systems of innovation: Towards a theory of innovation and interactive learning*. Pinter, London.
- Mahmood, I. P., & Mitchell, W. (2004). Two faces: Effects of business groups on innovation in emerging economies. *Management Science*, *50*(10), 1348-1365.
- Malerba, F. (1992). Learning by firms and incremental technical change. *The economic journal*, *102*(413), 845-859.
- Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica*, *65*(9-10), 1216-1223.
- Markard, J., & Truffer, B. (2006). Innovation processes in large technical systems: Market liberalization as a driver for radical change?. *Research policy*, *35*(5), 609-625.
- Markard, J., Raven, R., & Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. *Research policy*, *41*(6), 955-967.
- Meyer-Krahmer, F., & Reger, G. (1999). New perspectives on the innovation strategies of multinational enterprises: lessons for technology policy in Europe. *Research policy*, *28*(7), 751-776.
- Miller, D. J. (2006). Technological diversity, related diversification, and firm performance. *Strategic Management Journal*, *27*(7), 601-619.
- Miguelez, E., & Moreno, R. (2018). Relatedness, external linkages and regional innovation in Europe. *Regional studies*, *52*(5), 688-701.
- Mouri, N., Sarkar, M. B., & Frye, M. (2012). Alliance portfolios and shareholder value in post-IPO firms: The moderating roles of portfolio structure and firm-level uncertainty. *Journal of Business Venturing*, *27*(3), 355-371.
- Nan, D., Liu, F., & Ma, R. (2018). Effect of proximity on recombination innovation in R&D collaboration: an empirical analysis. *Technology Analysis & Strategic Management*, *30*(8), 921-934.
- Neffke, F. M., Otto, A., & Weyh, A. (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization*, *142*, 275-292.
- Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic geography*, *87*(3), 237-265.
- Neffke, F., Otto, A., & Hidalgo, C. (2016). The mobility of displaced workers: How the local industry mix affects job search strategies. *Papers in Evolutionary Economic Geography (PEEG) No*, *16*.
- Negro, S. O., Suurs, R. A., & Hekkert, M. P. (2008). The bumpy road of biomass gasification in the Netherlands: Explaining the rise and fall of an emerging innovation system. *Technological forecasting and social change*, *75*(1), 57-77.
- Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, *67*(3), 297-306.
- Nelson, R. R. (Ed.). (1993). *National innovation systems: a comparative analysis.* Oxford University Press on Demand.
- Nelson, R.R. and Winter, S.G. (1982). *An Evolutionary Theory of Economic Change.* Harvard University Press, Cambridge, MA.

- Nemet, G. F. (2009). Demand-pull, technology-push, and government-led incentives for non-incremental technical change. *Research policy*, *38*(5), 700-709.
- Nieto, M. J., & Santamaría, L. (2007). The importance of diverse collaborative networks for the novelty of product innovation. *Technovation*, *27*(6-7), 367-377.
- Nooteboom, B. (1994). Innovation and diffusion in small firms: theory and evidence. *Small business economics*, *6*, 327-347.
- Nooteboom, B. (2000). *Learning and innovation in organizations and economies*. OUP Oxford.
- Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & Van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research policy*, *36*(7), 1016-1034.
- Novelli, E. (2015). An examination of the antecedents and implications of patent scope. *Research Policy*, *44*(2), 493-507.
- Páez-Avilés, C., Van Rijnsoever, F. J., Juanola-Feliu, E., & Samitier, J. (2018). Multi-disciplinarity breeds diversity: the influence of innovation project characteristics on diversity creation in nanotechnology. *The Journal of Technology Transfer*, *43*, 458-481.
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- Phene, A., Fladmoe-Lindquist, K., & Marsh, L. (2006). Breakthrough innovations in the US biotechnology industry: the effects of technological space and geographic origin. *Strategic management journal*, *27*(4), 369-388.
- Prahalad, C. K., & Hamel, G. (1997). *The core competence of the corporation* (pp. 969-987). Physica-Verlag HD.
- Rallet, A., & Torre, A. (1999). Is geographical proximity necessary in the innovation networks in the era of global economy?. *GeoJournal*, *49*(4), 373-380.
- Ricardo, D. (1817). *On the Principles of Political Economy and Taxation.* John Murray, London. In: Sraffa, P., Ed., The Works and Correspondence of David Ricardo, Vol. 1, Cambridge University Press, Cambridge, 1951.
- Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy 19*(2), 165-174
- Sagar, A. D., & Van der Zwaan, B. (2006). Technological innovation in the energy sector: R&D, deployment, and learning-by-doing. *Energy policy*, *34*(17), 2601-2608.
- Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, *39*(8), 1051-1059.
- Schumpeter, J.A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credits, Interest, and the Business Cycle.* Transaction Publishers, Piscataway.
- Shannon, C. E., Weaver, W. (1964). *The Mathematical Theory Of Communication.* The University of Illinois Press.
- Shaw, A. T., & Gilly, J. P. (2000). On the analytical dimension of proximity dynamics. *Regional studies*, *34*(2), 169-180.
- Sidhu, J. S., Commandeur, H. R., & Volberda, H. W. (2007). The multifaceted nature of exploration and exploitation: Value of supply, demand, and spatial search for innovation. *Organization Science*, *18*(1), 20-38.
- Silverberg, G., & Verspagen, B. (2005). A percolation model of innovation in complex technology spaces. *Journal of Economic Dynamics and Control*, *29*(1-2), 225-244.
- Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. *Management science*, *56*(1), 41-56.
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society interface*, *4*(15), 707-719.
- Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, *44*(8), 1445-1461.
- Stuart, T. E., & Podolny, J. M. (1996). Local search and the evolution of technological capabilities. *Strategic management journal*, *17*(S1), 21-38.
- Su, H. N. (2022). How does distant collaboration influence R&D quality?. *Technology Analysis & Strategic Management*, *34*(7), 815-831.
- Tanner, A. N. (2014). Regional branching reconsidered: Emergence of the fuel cell industry in European regions. Economic Geography 90(4), 403-427
- Teece, D., & Pisano, G. (2003). *The dynamic capabilities of firms* (pp. 195-213). Springer Berlin Heidelberg.
- Tether, B. S. (2002). Who co-operates for innovation, and why: an empirical analysis. *Research policy*, *31*(6), 947-967.
- Torre, A., & Rallet, A. (2005). Proximity and localization. *Regional studies*, *39*(1), 47-59.
- Torre, A., & Wallet, F. (Eds.). (2014). *Regional development and proximity relations*. Edward Elgar Publishing.
- Unruh, G. C. (2000). Understanding carbon lock-in. *Energy policy*, *28*(12), 817-830.
- Utterback, J.M. (1996). *Mastering the Dynamics of Innovation.* Harvard Business Review Press
- Van Beers, C., & Zand, F. (2014). R&D cooperation, partner diversity, and innovation performance: an empirical analysis. *Journal of Product Innovation Management*, *31*(2), 292-312.
- Van den Bergh, J. C. (2008). Optimal diversity: increasing returns versus recombinant innovation. *Journal of Economic Behavior & Organization*, *68*(3-4), 565-580.

- Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).
- Van Rijnsoever, F. J., Van Den Berg, J., Koch, J., & Hekkert, M. P. (2015). Smart innovation policy: How network position and project composition affect the diversity of an emerging technology. *Research policy*, *44*(5), 1094-1107.
- Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research policy*, *45*(3), 707-723.
- Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research policy*, *45*(3), 707-723.
- Williamson, O. E. (1987). Transaction cost economics: The comparative contracting perspective. *Journal of economic behavior & organization*, *8*(4), 617-625.
- Winskel, M., Markusson, N., Jeffrey, H., Candelise, C., Dutton, G., Howarth, P., ... & Ward, D. (2014). Learning pathways for energy supply technologies: Bridging between innovation studies and learning rates. *Technological Forecasting and Social Change*, *81*, 96-114.
- Zhang, G., & Tang, C. (2018). How R&D partner diversity influences innovation performance: An empirical study in the nano-biopharmaceutical field. *Scientometrics*, *116*, 1487-1512.

Appendix A

Examples of project abstracts with high and low structural novelty

ID: *TKIW01006.*
Title: *Efficiency improvements by Lidar assistants,*
Techn. Novelty: *0.469*
Project Description:
Purpose: To reduce the Levelized Cost of wind energy by Reducing the Uncertainties of offshore wind farms Brief description: The project is built around testing, Evaluating and developing Lidar technology. Especially Lidar technology thatmakes use of the wind turbines in the wind farm as platforms. For instance power performce assessments using nacelle based lidars.

Example 1 high structural technological novelty, combining lidar technology with wind turbines (nacelles).

ID: *TEGB114015.*
Title: *Marktrijp ontwikkelen van varende warmte,*
Techn. Novelty: *0.48*
Project Description:
Purpose: This clever concept the Dutch water infrastructure is used to utilize waste heat efficiently for heating residential areas. The idea consists of heat storage in phase change material (PCM). Short description: The consortium is developing the concept, consists of DWA, ECN, HVC, Bronswerk Heat Transfer, Deen Shipping, INB-Group and the Port of Amsterdam. Interest in the concept is great. This is easily explained: the degrees of freedom (rest) to transport heat and maintain at a high level, are greatly expanded. Both distance and time is involved to a much lesser extent in the classical solution: the distance line. Moreover, the concept is easier to scale than a distance conduit in a start-up phase, with a positive result, the capital costs are lower and easier. The innovative part of the project lies both in the use of new PCMs suitable for higher temperatures as the scale. In comparison, the current storage / transport modality is made up of a 20-foot container, in which approximately 10 GJ of heat can be stored at a temperature around 57 ° C. scaling is necessary for a successful continuation of the concept, which reduces both the capital costs and operating costs. Reason: DWA has launched the concept of sailing heat. The innovative part of the concept is, inter alia, in storing heat at a high temperature level (80 to 90 ° C), whether or not in combination with the transport of heat by ship. Result: Both the local government (counties) as heat production, industry and waste treatment is interest in the concept. The industry is also the static application for the storage of important process heat: the improvement of the heat balance, thereby increasing the energy efficiency. Striking is the width in which applications are also possible. Industrial partners rather opt for higher temperatures; users can do with lower temperatures in the built environment. This gives its own dynamics.

Example 2 high structural technological novelty, heat transport on water for heating residential areas

TKIEI01002SB: Cluster Sustainable Business Models
Creating new business models where energy savings and energy efficiency are realized through new (intersectoral) partnerships, using existing innovative technologies where the economy of chain is one of the drivers. The business environment is changing rapidly due, among other things, to rising energy prices, resource scarcity and social pressure for sustainability and local initiatives. This creates the opportunity for developing new business models that make economically feasible energy savings (across a chain). The condition for this is to actually connect parties in the chain and reduce risks for all parties and to develop earnings models.

Example 1 low structural technological novelty: new business models for energy efficiency

ID: *TKIG01019.*
Title: *Fractures from basin to well scale: num. stress-strain model of fracture networks,*
Techn. Novelty: *0.438*
Project Description:
Objective: The objective of this project is to obtain knowledge about the mechanical behavior of a shale gas reservoir on the increase of the liquid pressure, wherein the effective voltage is reduced around the borehole. This should take into account existing fractures (and other sedimentary / structural heterogeneities) The research will include predictions on how the proceeds of the reservoir can be increased by breaking rocks in small to medium scale and how the risks of triggering larger fractures can be limited. Brief description: This project will develop geomechanical models at multiple length scales (104-101m) to determine the tension in relevant parts of Dutch schaliebekkens (Variscisch promontory, London-Brabant Massif). The large fractures are introduced in a deterministic manner, small fractures and cracks are introduced optionally stochastic. Then, i) the tension fields must be correctly predicted in the area, and ii) investigating the possible reactivation. The geometric characteristics of smaller cracks can be predicted based on analyzes of rock layers on the surface that are similar to the reservoir rocks at depth. Results: Results at the end of 2013 (the pilot year): A three-dimensional geometric model of a test site in the Dutch subsurface (selected in coordination with related projects and sponsors), which describes the mechanical structure of rock layers, including medium to large fractures. Evaluation of the current state of stress in the respective volume on a scale of 102-104m. With the aid of numerical computer modeling the deviations are determined from the stress around large-scale natural fracture system, and predictions are made about the possibility of causing cracks on a smaller scale. Forecasts of change in the voltage orientation help around major fractures in determining critical loads and the risk of reactivation. In year 1, the voltage patterns are determined for heterogeneities at multiple scales and under natural loads. In years 2 and 3 will be given to intermediate scales from medium-sized cracks to cracks around the hole and to the reduction in effective printing fluid injection.

Example 2 low structural technological novelty: shale gas extraction research

ID: *TKIG01025.*
Title: *Fracture initiation fracture growth, fluid flow and particle transport,*
Techn. Novelty: *0.426*
Project Description:
Purpose: Predicting the geometry of the fracture network, Which will develop in the areas experiencing fracturing fluids are once injected, will Improve our ability to predict the characteristics and behavior tough or gas reservoirs. In this we developed a project, a numerical model to predict the fracturing process in hydraulic fracturing situations. Only an accurate prediction of the hydraulic fracture network makes it possible to estimate permeability changes, the degree of fluid penetration into the rock matrix and, Eventually, investigate Their impact on matrix-to-fracture gas transfer rates.

Example 3 low structural technological novelty: research hydraulic fracturing

Examples of projects with high functional novelty

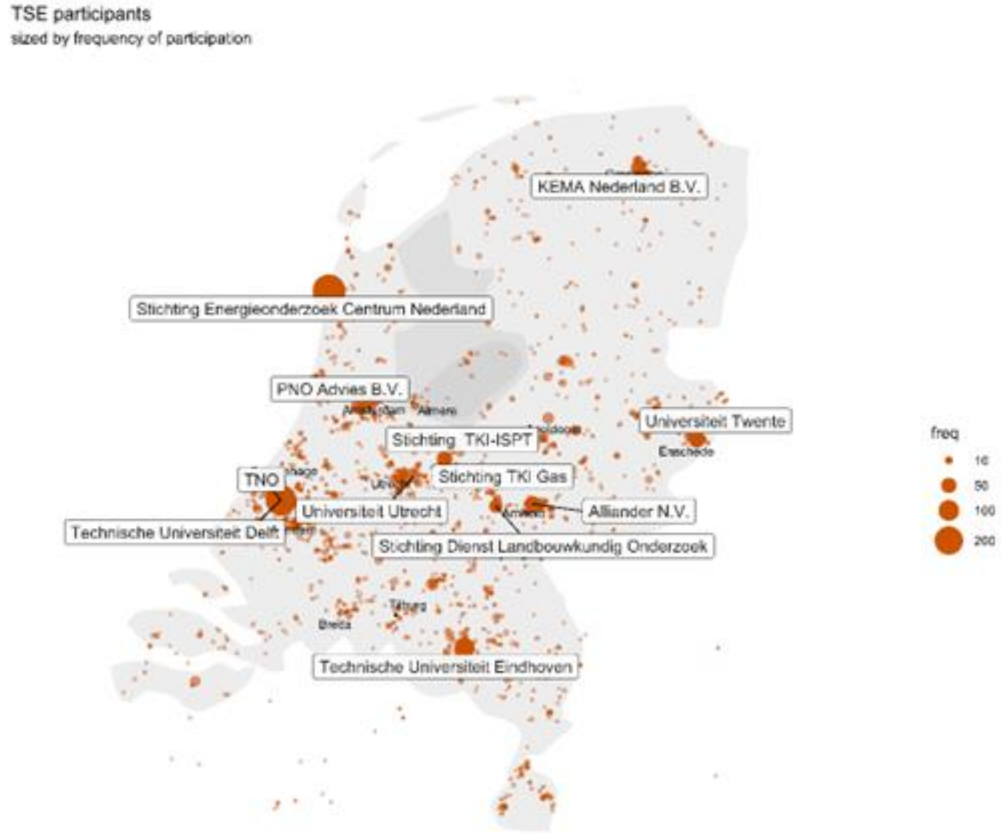
- TKIZ01012, Roll-to-roll organics for PV, IEA Category C1. Solar Energy: The use of inkjet technology to improve the performance of organic photovoltaic modules[7]
- DEI1160026, Energiebesparing door MCFA productie uit voedselresten, IEA Category A1. Energy efficiency: industry. The production of Medium-chain fatty acids (MFCA) from food scraps, rather than using conventional petrochemical based manufacturing[8].
- TESG113001, Smart Grid in Balans, IEA category F2. Electricity transmission and distribution: Creating a link between the supply of decentralized sustainable energy sources and the demand for sustainable energy from electric transport[9].

Examples of projects with low functional novelty

- TEZG114006, Intelligente elektronica voor (fotovoltaïsche) zonnesystemen, IEA Category C1. Solar Energy: Develop and validate an electronic circuit that can be integrated into photovoltaic (PV) modules to improve energy harvesting[10].
- TEBG113001, Demonstration of a treatment plant utilizing anammox bacteria. IEA Category C4, Biofuels. This technology can potentially be applied to other (industrial) digesters in the future[11]

[7] https://projecten.topsectorenergie.nl/projecten/r2ro4pv-roll-to-roll-organics-for-pv-11956
[8] https://data.rvo.nl/subsidies-regelingen/projecten/energiebesparing-door-mcfa-productie-uit-voedselresten-plaats-van-olie
[9] https://projecten.topsectorenergie.nl/projecten/smart-grid-in-balans-16588
[10] https://projecten.topsectorenergie.nl/projecten/intelligente-elektronica-voor-fotovoltaische-zonnesystemen-in-gebouwde-omgeving-18536
[11] https://projecten.topsectorenergie.nl/storage/app/uploads/public/5b7/acb/5c5/5b7acb5c56199091859691.pdf

Appendix B: Geographic locations of Dutch TSE participants



Appendix C

Table 13 Regression results using expert innovativeness rating as dependent variable.

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	-11.37	[-64.98, 42.25]						
year	0.01	[-0.02, 0.03]	0.02	[-0.09, 0.14]	.00	[-.00, .01]	-.00	
funding	0.00*	[0.00, 0.00]	0.14	[0.02, 0.25]	.02	[-.01, .05]	.12*	
trl_level	-0.05	[-0.26, 0.15]	-0.19	[-0.94, 0.55]	.00	[-.01, .01]	.05	
trl_level_squared	0.01	[-0.02, 0.04]	0.25	[-0.50, 1.00]	.00	[-.01, .01]	.06	
tech_proximity	-0.05	[-0.18, 0.08]	-0.05	[-0.16, 0.07]	.00	[-.01, .01]	-.05	
log_distance_km	-0.00	[-0.03, 0.03]	-0.00	[-0.12, 0.11]	.00	[-.00, .00]	.02	
org_balance	0.16**	[0.08, 0.25]	0.23	[0.12, 0.35]	.05	[.00, .09]	.23**	
								R^2 = .079** 95% CI[.01,.12]

b represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

Appendix D, training custom word vectors

We apply GloVe (Global Vectors for Word Representation) to learn a word vector space on a training set, consisting of 2,356,000 patents. In the vector space energy related words are presented as n-dimensional vectors, representing the energy technology space. To arrive at an optimal number of dimensions given the data, different candidate numbers of dimensions were tested on a validation task. It was found that a model with 200 dimensions resulted in the best performance on the validation tasks. However, we only achieve an f1 of .25 on the accuracy test, which is far below the pretrained GloVe model.

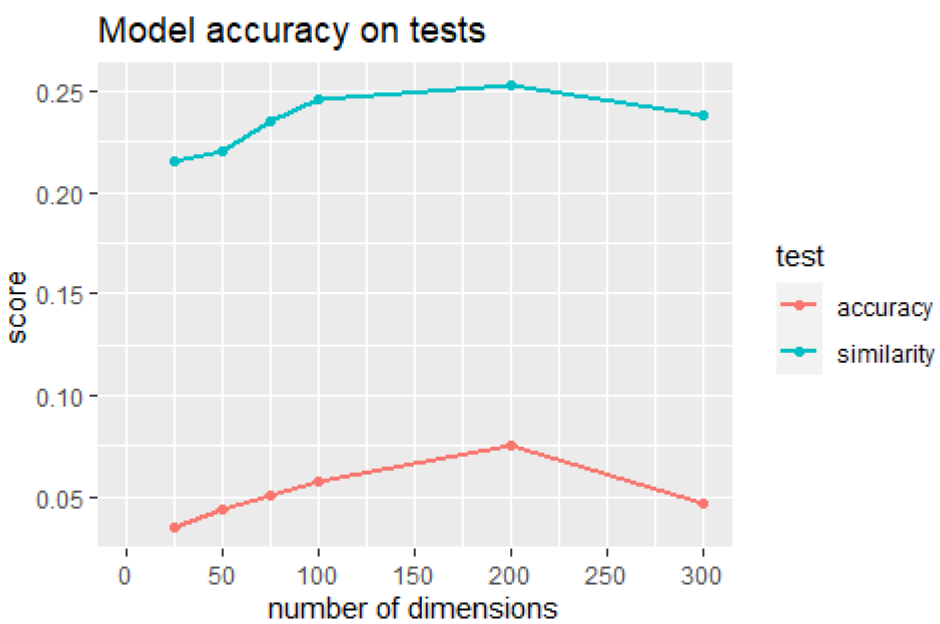
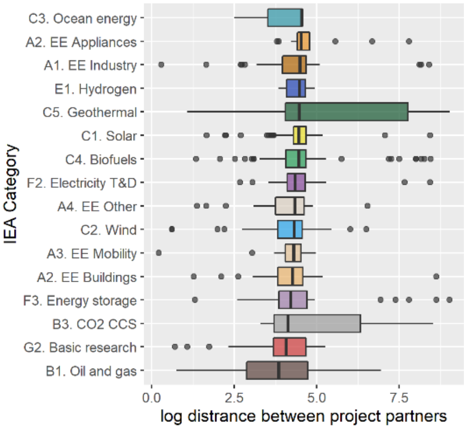
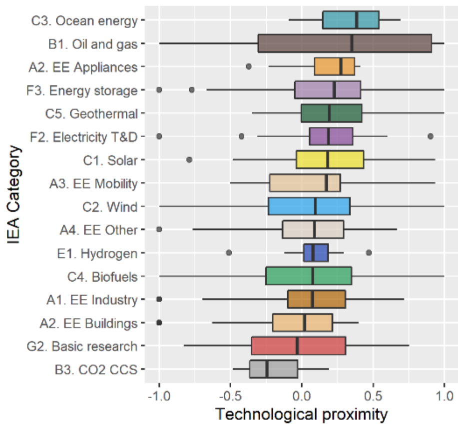
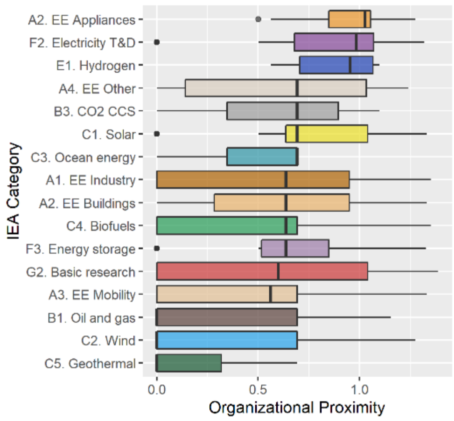


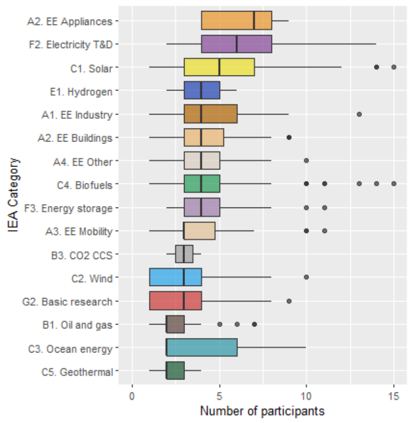
Figure 13: Validation tasks (analogy test and similarity test)

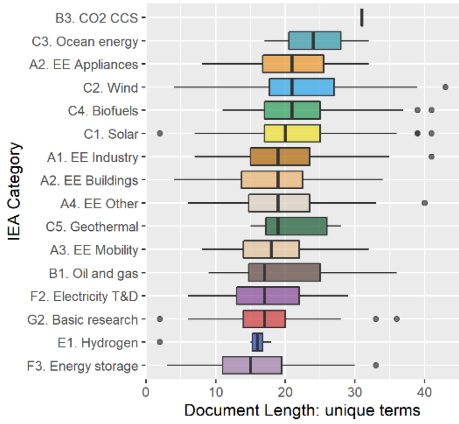
Appendix E

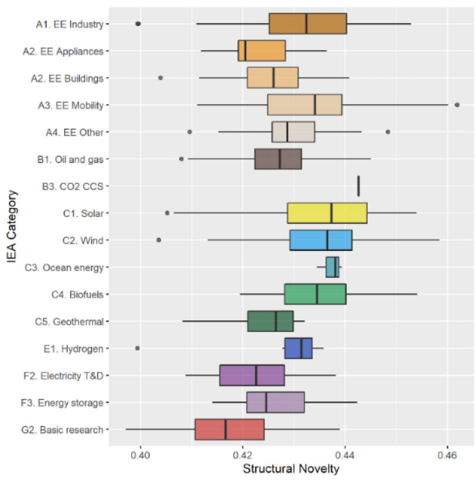












Grouped by IEA Category we find that solar, wind and ocean energy projects exhibit on average the most structural novelty. Electronic appliances, basic research and electricity transmission & distribution projects are on average least novel.