



**HAL**  
open science

# Impact of knowledge search practices on the originality of inventions: A study in the oil & gas industry through dynamic patent analysis

Quentin Plantec, Pascal Le Masson, Benoît Weil

## ► To cite this version:

Quentin Plantec, Pascal Le Masson, Benoît Weil. Impact of knowledge search practices on the originality of inventions: A study in the oil & gas industry through dynamic patent analysis. *Technological Forecasting and Social Change*, 2021, 168, pp.120782. 10.1016/j.techfore.2021.120782 . halshs-03203124

**HAL Id: halshs-03203124**

**<https://shs.hal.science/halshs-03203124>**

Submitted on 20 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# IMPACT OF KNOWLEDGE SEARCH PRACTICES ON THE ORIGINALITY OF INVENTIONS: A STUDY IN THE OIL & GAS INDUSTRY THROUGH DYNAMIC PATENT ANALYSIS

Quentin Plantec<sup>1,2</sup>, Pascal Le Masson<sup>1</sup> & Benoît Weil<sup>1</sup>

1. MINES ParisTech, PSL University, Centre for Management Science (CGS), i3 UMR CNRS, 60 Boulevard St Michel, 75006 Paris, France<sup>1</sup>

2. Institut National de la Propriété Industrielle (INPI), 15 rue des Minimes, 92677 Courbevoie CEDEX, France

## ABSTRACT:

*The paper suggests a new taxonomy of knowledge search modes to describe the creative process of new invention design, in particular how firms combine knowledge components from their own knowledge base—taking into account both the components and the structures of knowledge bases—with those from newly acquired or newly internally developed. Using network theory techniques, we defined four knowledge search modes: (1) refinement, (2) clustering, (3) absorption and (4) recombination. We conducted an exploratory study on the oil & gas industry, reviewing 50,776 utility patents filed by 16 major firms between 1989 and 2016. The results showed, first, that firms relied to varying extents on different knowledge search modes in their invention design processes. Second, reviewing the technological originality of the designed inventions showed that simply absorbing new knowledge components, without major changes in knowledge base structure, was associated with low technological originality, but constituted one of the main knowledge search modes used by the analyzed firms. In contrast, major changes in knowledge base structure favored technological originality, with or without new knowledge components, but were nevertheless the least used mode. Understanding organizational learning practices associated with the phenomena described here can foster innovation performance in firms.*

## HIGHLIGHTS:

- Oil & gas firms are heterogeneous in their knowledge search practices for designing inventions.
- Firm's knowledge expertise should be bridged to design original inventions.
- New knowledge not bridged with firm's expertise negatively affect originality.

**KEYWORDS:** Knowledge search, patent analytics, oil & gas industry, technological originality, knowledge base

**RUNNING TITLE:** Knowledge search and invention originality

**ACKNOWLEDGMENT:** We would like to thanks B. Cabanes (MINES ParisTech – PSL University), C. Sternberger (INPI), N. Sennequier (INPI), F. Caillaud (WIPO), C. Vahlibay (MINES ParisTech—PSL University), G. de Rassenfosse (EPFL), M. Wustmans (Bonn University), M. Chebre, J. Luo (IDC), S. Sarica (IDC) and the participants of the Data & Algorithms for Science, Technology and Innovation Studies Conference organized by EPO and KU Leuven for their useful insights in relation to this research project.

**FUNDING:** This work was supported by the Institut National de la Propriété Industrielle (INPI) [collaborative Ph.D. (CIFRE scheme) grant number 2018/0255] and the Chair of Design Theory and Methods for Innovation (DTMI) of MINES ParisTech – PSL University.

**DECLARATION OF INTEREST:** The Chair of Design Theory and Methods for Innovation (DTMI) of MINES ParisTech – PSL University is financed by more than ten different companies and institutions. One of them was part of the sample of the present study.

– WORKING PAPER VERSION –

To cite: “*Impact of knowledge search practices on the originality of inventions: a study in the oil & gas industry through dynamic patent analysis*”. Plantec, Q., Le Masson, P., Weil, B., *Technological Forecasting and Social Change*, vol. 168, 2021. <https://doi.org/10.1016/j.techfore.2021.120782>.

# 1. INTRODUCTION

Technological novelty is an important driver of innovation (Arthur, 2007; Fleming and Sorenson, 2004), which is presumed to have a large impact on organization competitiveness and market position (Cooper and Schendel, 1976; Henderson and Clark, 1990). However, innovation is a complex process with high uncertainty across different stages, from the front end to the launch of the newly developed product and its subsequent social adoption and market diffusion (Fleming, 2001). There is a crucial need to better understand a particular part of the latter process: the knowledge search supporting the design of new inventions (Arts and Fleming, 2016). In particular, a firm-level analysis must be conducted for this knowledge search process because inventors of firms are embedded in particular knowledge networks and have distinct design capabilities (Ahuja and Lampert, 2001; Nootboom et al., 2007).

Various streams of research on innovation have highlighted the importance of knowledge search activities (e.g., Chesbrough, 2003; Cohen and Levinthal, 1990; Hatchuel and Weil, 2003; Nelson and Winter, 1982; Nonaka, 1994). We rely here on the literature on recombinant innovation (Ahuja and Lampert, 2001; Fleming, 2001; Fleming and Sorenson, 2001; Rosenkopf and Nerkar, 2001): knowledge search activities aimed at selecting and combining knowledge components to design new inventions. As described by March (1991), firms combine knowledge components through an exploitative approach based on the firm's current knowledge and associated close opportunities (i.e., local search) and/or through an explorative approach based on new and distant possibilities (i.e., distant search). The set of knowledge components mastered by the firm is part of its knowledge base (Grant, 1996), **and these components are** referred to as **knowledge base components** here. The knowledge of the firm is embedded in its combinative capabilities because knowledge combination is involved in a recombinant innovation perspective as well (Kogut and Zander, 1992). Hence, the knowledge base has a particular structure; that is, different couplings exist among all knowledge components mastered by the firm based on the ability of the firm to combine them (Le Masson et al., 2010a; Yayavaram and Ahuja, 2008).

To what extent(s) firms need to rely on different combinations of newly sourced or already mastered knowledge components to support the design of original inventions is a key question for both innovation management and the analysis of technological dynamics. To help further the understanding of this issue, scholars have analyzed knowledge search practices from the perspective of the global technological landscape, exploring to what extent inventors were combining already known or new knowledge components and what the effects were of those practices on the value of the designed inventions (Arts and Fleming, 2016; Fleming, 2001; Lobo and Strumsky, 2019; Strumsky and Lobo, 2015; Verhoeven et al., 2016). Since the 1990s, inventors have apparently been extensively relying on the refinement of the existing combinations of known knowledge (Strumsky and Lobo, 2015), while at the same time relying on very new knowledge have been associated with a higher probability of achieving a breakthrough invention (Arts and Fleming, 2016; Fleming, 2001), in particular when based on new scientific knowledge (Choi et al., 2018; Verhoeven et al., 2016). In addition, a few scholars have successfully reviewed the role of changes in a firm's knowledge base following knowledge search activities (integration of new components and/or new combinations) on the firm's innovation performances (Yayavaram and Ahuja, 2008; Yayavaram and Chen, 2015). They have shown that changes in the combinations between existing knowledge components of firms hurt innovation performance, while new combinations between new and established knowledge components favor innovation performance.

**Our present contribution aims to investigate the plurality of knowledge search practices undertake by firms and their impacts on the technological originality of**

**designed inventions by taking into account the firm's knowledge base (components and structures).** Technological originality is marked by the significant difference between combinations of underlying knowledge components embedded in a particular invention and the predominant design. In that sense, we aim at bridging the literature on firms' knowledge base and that on inventors' knowledge search practices. Indeed, our approach acknowledges that inventors are embedded in firms that support their innovation efforts through distinctive innovation capabilities, team management practices, and exploration practices (Lawson and Samson, 2001; Le Masson et al., 2010b). The combination of knowledge components examined by an inventor of the firm to design an invention is associated with the firm's knowledge base.

In this study, we develop a taxonomy of four knowledge search modes that are used to support the design of novel inventions depending on the firm's knowledge base structure: (1) refinement, (2) clustering, (3) absorption, and (4) recomposition. Our taxonomy was implemented through an exploratory study of 16 major oil & gas firms involving a review of their patent portfolios between 1989 and 2016, comprising 50,776 utility patent applications. We show that, first, the knowledge search modes adopted by firms vary among time and across the firms in the sample. Second, however, refinement was the most used knowledge search mode in the oil & gas industry, but led to inventions with low technological originality. Third, the absorption mode, the second-most used, which focuses on the integration of new knowledge components without major changes in the firm's previous knowledge base structure, scored poorly in terms of technological originality. In contrast, the recomposition knowledge search mode, a design process that generates major changes in firms' knowledge base structure (with or without integration of new knowledge components), conduces to highly original inventions, but was nevertheless the least adopted among the analyzed firms, indicating a major organizational learning challenge. Fourth, firms somewhat rely on a clustering mode as its technological originality score is better than that of the absorption mode. Finally, the research opens up promising new research paths for the innovation field, emerging from the implication that technological originality, when integrating new knowledge components, depends on implied changes in the firm's knowledge base structure. This also has consequences for organizational learning and innovation design, as it calls for taking into account the dual status of knowledge component distance, that is, considering it both from the firm's and the firm's inventor's points of view.

## 2. RESEARCH QUESTIONS

There is a consensus in the innovation literature that firms need to develop skills in both internal knowledge development and external knowledge sourcing (Helfat, 1994). The activities related to gathering knowledge components—either gained from internal development or sourced externally—to develop new inventions are referred to as knowledge search practices.<sup>2</sup> They are closely linked to the firm's knowledge base (Ahuja and Katila, 2001). When analyzing both firms' knowledge base structures and components, it appears that there is a plurality of knowledge search practices that can be undertaken by firms depending on the extent to which they are relying on local or distant knowledge, respectively (Yayavaram and Chen, 2015). We address here the question of how to organize and choose among a plurality of knowledge search practices to efficiently design new, original inventions. The following research questions are addressed using patent analytics techniques, as a firm's

---

<sup>2</sup> Katila (2002) used "innovation search," but we refer here to "knowledge search," a term more commonly used in the literature.

patent applications are a proxy for the firm's knowledge base (Jaffe, 1986). Our research questions are then the following (analytical framework and research questions are synthesized in Figure 1):

- QR1: *How are knowledge search practices undertaken by firms to design novel inventions related to both their previous knowledge base—components and structures—and the sourcing of new external knowledge?*
- QR2: *How are knowledge search practices related to the technological originality of those inventions?*

where

**Firm's knowledge search practices** denote the different means employed by a firm to select and then combine diverse sources of knowledge components from the knowledge space in order to design a new invention.

**A firm's knowledge base** includes both (1) **the knowledge components** comprising expertise, knowledge and skills already acquired and mastered by the company and (2) **the structure** of the knowledge components describing interactions among these components. The structure describes how they have been combined together by the firm over time in order for improving design products, processes, or artefacts or for Research & Development purposes. A combination of two components is then considered a **knowledge coupling**.

and

**Technological originality of an invention:** there are many ways to measure technological originality of an invention. Here, it refers to the degree of difference between the underlying knowledge couplings embedded in a particular invention and a predominant design, which is conceptualized as a long-term and multi-industry technological knowledge base structure.

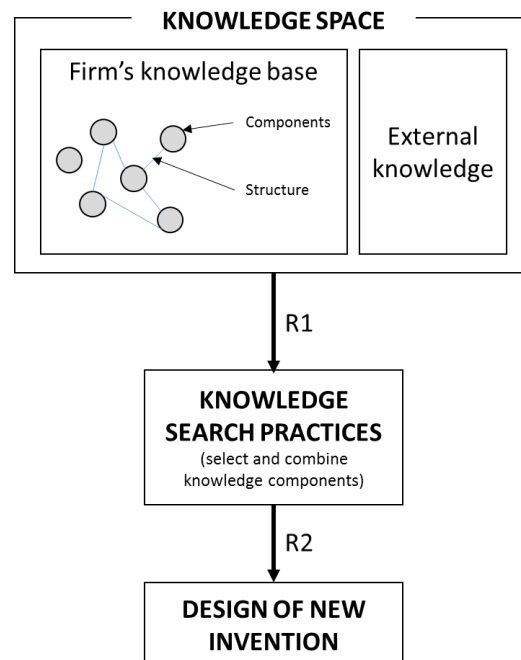


Figure 1 - Analytical framework and research questions

### 3. LITERATURE REVIEW

As we are focusing on patent data as a proxy for firm's knowledge base, the following literature review aims to define the theoretical background with a particular emphasis on the patent analytics techniques that have been developed in the research field.

#### 3.1. Knowledge search practices: (re)combination and the design of new inventions

Various streams of research on innovation have highlighted the importance of knowledge search activities, in areas such as organizational learning (e.g., Cohen and Levinthal, 1990; Nonaka, 1994), design science (e.g., Hatchuel and Weil, 2009), evolutionary theory (e.g., Nelson and Winter, 1982) or open innovation (Chesbrough, 2003; West et al., 2014). To design new inventions, scholars mainly refers to the problem-solving logic inspired by the seminal Nelson and Winter (1982) book, under which firms' knowledge search activities aim at solving problems by combining knowledge components to create new products.

Hence, innovation scholars have mainly operated under the assumption that combining existing technological capabilities or knowledge components in novel ways (potentially including new knowledge for new combinations) is one of the principal sources of inventive novelty (Dosi, 1982; Fleming, 2001; Fleming and Sorenson, 2001; Hargadon and Sutton, 1997; Henderson and Clark, 1990; Schumpeter, 1934; Teece, 1996). Fleming and Sorenson (2001) gave a few examples of combinatorial inventions: *“one might think of the automobile as a combination of the bicycle, the horse carriage, and the internal combustion engine. The steam ship can be characterized as combining the boat with steam power. Similarly, one might consider the microprocessor to be a conjunction of a computer's central processing unit with integrated circuit fabrication processes”* (Fleming and Sorenson, 2001, p. 1020). Three main ways of combining knowledge can be identified: (1) combining new knowledge only, (2) designing new configurations of previously combined knowledge, and (3) mixing new knowledge with previously combined knowledge (Fleming, 2001; Yayavaram and Chen, 2015). In terms of innovation management, this implies that, first, firms need to develop combinative capabilities in order to create new configurations without becoming trapped in complexity issues (Fleming, 2001; Kogut and Zander, 1992). Second, firms have to develop organizational learning processes (e.g., routines) that foster the availability of the internal expertise that could be required in the design of new inventions (Ahuja and Lampert, 2001; Nonaka, 1994). Third, firms need to support internally developed new knowledge or to be involved in inside-in open innovation activities (Gassmann and Enkel, 2004) to access new knowledge components.

Thus, a firm is considered to be adopting local search practices when it is designing new inventions, in which majority of the technological content is similar to its prior search activities (Stuart and Podolny, 1996). This system was the main practice for much of the 20<sup>th</sup> century, while corporate research and development (R&D) labs were embedded in vertically integrated infrastructure in firms such as AT&T, DuPont, or IBM (Arora et al., 2018; Buderl, 2000; West et al., 2014). Firms also undertake distant search practices, in which knowledge created by other actors, such as competitors, consumers, or academic partners, is accessed and mastered by the firm to design new inventions (Ahuja and Lampert, 2001; Rosenkopf and Nerkar, 2001). Firms' internal practices can foster distant knowledge search by reviewing competitors' published patents, attending scientific or industry conferences, reviewing scientific articles, engaging in partnerships, etc. The literature on open innovation (Chesbrough, 2003) has opened new paths for in-depth analysis of external knowledge search processes (West and Bogers, 2014); in particular, many business approaches can be used to foster open innovation, such as outsourcing of R&D activities, joint development, purchase of external technology, or mergers and acquisitions (Cammarano et al., 2017).

Based on these theoretical assumptions, scholars have developed techniques based on patent analytics to explore how inventions are designed. Those models mainly use Patent Classification Codes such as the International Patent Classification (IPC) system as a proxy for knowledge components. For example, Lobo and Strumsky (2019) and Strumsky and Lobo (2015) used four categories to describe novel inventions: (1) **origination** if all the IPC codes used in a given patent have never been used before, (2) **novel combination** if at least one pair of IPC codes in a given patent contains a code that has never been used before, (3) **combination** if at least one pair of IPC codes has never been used before in a given patent, and finally, (4) **refinement** for cases where all IPC codes have already been used before in other patents. They showed that from 1980 to 2014 for United States Patent and Trademark Office (USPTO) patents, combinations and refinements accounted for 47% and 52% of all patents respectively, while origination and novel combinations were very rare. Verhoeven et al. (2016) used another taxonomy: (1) **novelty in recombination**, when a patent contains at least one pair of IPC codes that was previously unconnected, (2) **novelty in technological origins**, when a patent contains at least one pair of IPC codes between the focal patent and the patent cited as literature (backward citation) that was previously unconnected, and (3) **novelty in scientific knowledge**, based on the pair of IPC and Web of Science category classifications of scientific articles cited as literature in the patent (i.e., Non-Patent Literature). By using a sample of European Patent Office (EPO), USPTO and World Intellectual Property Organization (WIPO) patent families, they showed that 7% of patent families scored on novelty in recombination, 22% scored on novelty in technological origins, and 11% of patent families that cited at least one scientific article scored on novelty in scientific knowledge. Overall, 25% of all patent families scored on at least one novelty criteria.

These valuable techniques help us better understand the nature of the knowledge search practices used by inventors to design new inventions and their occurrence in the global technological landscape. Furthermore, still based on patent analytics, scholars have also developed more applied research conducting analyses at firm level, in order to help a given firm identify what new knowledge components could be combined with their established knowledge to design new inventions (e.g., Nakamura et al., 2015; Sarica et al., 2019). Thus, the design of new inventions is closely related to a firm's knowledge base, whose structure matters in this regard.

### **3.2. Firm's knowledge base—components and structures—and its interface with knowledge search practices**

The firm's knowledge base describes what the firm knows (Grant, 1996; Jaffe, 1986; Kogut and Zander, 1992). Knowledge base theory has established that knowledge components are the preeminent resource of innovation activities of firms (Grant, 1996). Nevertheless, not only the knowledge components themselves matter: the structure of the knowledge base is also a key element (Kogut and Zander, 1992). Hence, the knowledge base "*can be characterized by the set of knowledge elements that it possesses and the relationships that it has forged between the knowledge domains to which these elements belong*" (Yayavaram and Chen, 2015, pp. 377–378). As a result, firms' knowledge search practices are mainly linked to the firm's knowledge base: local knowledge corresponds to the firm's knowledge base at time  $t$ , while accessed distant knowledge components fuel the firm's knowledge base and are integrated at time  $t+1$ , as well as internally developed new knowledge generated through the local space (Kogut and Zander, 1992; Yayavaram and Ahuja, 2008; Yayavaram and Chen, 2015). It has to be highlighted that there are many possibilities for combining ancient local

knowledge, newly developed local knowledge, and distant knowledge in order to design a novel invention.

Developed patent techniques also exist to explore firms' knowledge base components and structures. At the adopted starting point, scholars used Patent Classification Codes to reveal the knowledge components used by a given firm. For example, Fleming and Sorenson (2001) used USPTO subclass references for all patents filed by a given firm to compute knowledge components of the firm's knowledge base. They also explored couplings between knowledge components in the knowledge base, but from a global market perspective and not from that of the focal firm. Yayavaram and Ahuja (2008) and Yayavaram and Chen (2015) provided a more advanced representation of a firm's knowledge base by focusing on knowledge components and couplings at a firm-level analysis. "[It] suggests that for several reasons it is beneficial to conceptualize knowledge bases as networks of knowledge elements in which even the ties between knowledge elements are important, rather than as simply sets of individual elements." (Yayavaram and Ahuja, 2008, pp. 357–358). Yayavaram and Ahuja (2008) focused on the link between firms' knowledge structure and the usefulness of inventions using network theory techniques, revealing a firm's knowledge base by computing all pairs of knowledge components sourced from the firm's patent applications. First, they showed that there is an inverse-U-shape between continuums of structures—from fully decomposed to an integrated knowledge base—and inventions' usefulness as measured through patent citations. Second, they highlighted that firms are pursuing two types of strategies: (1) adding new knowledge components and (2) adding new combinations of existing knowledge components (i.e., new couplings). Furthermore, Yayavaram and Chen (2015) explored the extent to which changes in knowledge component couplings between (1) familiar knowledge components only and (2) both new knowledge and familiar knowledge components affects innovation performance, using knowledge base complexity as a moderating variable. By reviewing financial and patent data for 1,750 firms between 1976 and 2004, they showed that changes in coupling of knowledge with which the firm is familiar (i.e., local knowledge) hurt the firm's innovation performance, while new couplings, using both local and distant knowledge components, increase the firm's innovation performance. Drawing on the literature in architectural innovation (Henderson and Clark, 1990), they highlighted that changing couplings between local knowledge components in highly complex environments reflects that the firm has overcome interdependency issues and uncertainty regarding which couplings are valuable. This phenomenon then leads to improved innovation performance.

It has to be noted that the usage of network theory techniques with patent data, as in Ahuja's (2008) approach using those data as a proxy knowledge base, is a familiar approach in the innovation research field. Nevertheless, those techniques have been extensively used mainly for other purposes: to build a global map of technological proximity (see Alstott et al., 2017a; Leydesdorff et al., 2014 for extended discussions), as well as to represent knowledge dynamics through inventor networks (e.g., Fleming and Frenken, 2007), partner networks (e.g., Grigoriou and Rothaermel, 2017), or citation networks (Li et al., 2019). In contrast, to our knowledge, very few studies have combined the usage of Patent Classification Codes and network theories tools in order to reveal firms' knowledge structure in the management science field, with the only known exceptions being Yayavaram and Ahuja (2008) and Yayavaram and Chen (2015). Although scholars have computed global maps of technological proximity based on Patent Classification Codes and used a few focal examples to illustrate their insights (cite examples), their research has not focused on individual firms. Furthermore, while reviewing IPC lists at a firm level is a common indicator of firm competitiveness (e.g., Jaffe, 1986)—in particular in the economics field—it has not been extensively used supported by network theory or to focus on invention design and scientific or technical evaluation. We also note that in another context, Cabanes (2017) used advanced network theories (matroid theory) techniques



to model firm expertise dynamics based on knowledge independencies. Nevertheless, those techniques relied on a qualitative data coding scheme that is not applicable to large-scale quantitative patent data.

### **3.3. Sourcing knowledge locally or distantly and the quality of invention**

As firms rely on different knowledge search modes to source and combine knowledge components in order to design new inventions, with various implications and dependencies regarding firm's knowledge base, one key question is: to what extent does a given firm need to rely on its previous stock of knowledge to foster the design of new inventions (Katila, 2002), that is, to what extent does it hold that "*old is gold*" (Nerkar, 2003)?

A wide stream of research argues that radical innovation necessarily presupposes the utilization of very recent knowledge (Henderson and Clark, 1990; Katila, 2002; O'Connor, 2008) and knowledge from different industries and technological domains (Dahlin and Behrens, 2005; Dosi, 1982; Katila and Ahuja, 2002; Nootboom et al., 2007). In principle, such knowledge has not yet been integrated into the firm's knowledge base. By reviewing novelty through patent citations, Ahuja and Lampert (2001) effectively showed that having no backward citations, which can be interpreted as only using new knowledge components, is associated with more radical invention. Furthermore, Arts and Fleming (2016), relying on semantic analysis of patent data, showed that inventors who change fields are more likely to create novel patents due to their exposure to new knowledge components. In addition, drawing on state-of-the-art new scientific knowledge is also associated with the exploration of completely new innovation paths, such as the use of spintronic theories developed by 2007 Nobel laureate Albert Fert for semiconductors and the current industrial development on graphene by Nobel laureate Andre Geim (e.g., Fleming and Sorenson, 2004; Hatchuel et al., 2013). By measuring scientific article novelty and utilization in patent literature, Veugelers and Wang (2019) showed that utilizing more original scientific articles conduces to more novel patents (through patent citation evaluation). Finally, firms that build on newly developed knowledge components more frequently are able to better predict future technological advances and thereby design more novel inventions (Cohen and Levinthal, 1989). The frequent exposure to newly developed knowledge helps firms to sustain their cognitive capabilities, thereby helping future innovation (March, 1991).

Although relying on external knowledge is positively related to the originality of the design of a given invention, extensive reliance on local knowledge has been proven to be detrimental to the global innovation process of firms. First, prior learning and existing paradigms can cause learning myopia or learning traps (Levinthal and March, 1993) and constrain the direction of search due to cognitive path dependence, specific historical pathways, or fixation effects (Agogué et al., 2014; Kaplan and Tripsas, 2008; Sydow et al., 2009). Nevertheless, it has been underlined by many researchers that firms should exploit and build upon their knowledge and expertise (i.e., their knowledge base) to innovate; indeed, as technological progress is cumulative, inventors inevitably need to draw on their prior knowledge and expertise (Kline and Rosenberg, 1986). Using knowledge components which the firm is familiar with enhances the reliability of the design and its uniqueness, in particular if the firm has a distinctive knowledge base (Katila, 2002; March, 1991). Following a combinatorial logic, a larger knowledge base evidently give more rooms for combinatorial possibilities, but it can also enhance complexity due to interdependency of knowledge components (Fleming, 2001; Yayavaram and Ahuja, 2008).

March (1991) characterizes the essential problem as a choice between exploiting known knowledge or exploring new and distant possibilities. Literature on “innovation ambidexterity” has drawn on those two concepts, advocating for simultaneously pursuing both exploration and exploitation (e.g., Andriopoulos and Lewis, 2010; O’Reilly and Tushman, 2008; Raisch et al., 2009), but mainly through pursuing different projects on similar timeframes. There is also room for the combination of both new and old knowledge components at an invention level and not at a portfolio level (Hatchuel and Weil, 2009; Le Masson et al., 2010a; Strumsky and Lobo, 2015; Yayavaram and Chen, 2015).

### 3.4. Research gap

To design novel inventions, firms need to combine knowledge components based on a continuum of search practices, depending on the extent of their reliance on familiar knowledge components. In our approach, we are bridging the literature on inventor’s knowledge search and firm’s knowledge base, focusing solely on invention originality but taking into account that the inventor is embedded in a given firm with distinct innovation practices and knowledge base (Le Masson et al., 2010a). The combination of knowledge searched for to design a given invention is then related to the components, the structure of the knowledge base, and in particular, the distance between knowledge components in the knowledge base. Thus, what would previously have been considered “local” knowledge from a firm’s perspective is not necessarily local for a given inventor in that firm.

To our knowledge, we are the first to provide a **formal model describing in-depth how** firms are designing new inventions depending on their knowledge base structure and the combination of local and distant knowledge components. As noted by Yayavaram and Ahuja (2008): “*The problem of structuring organizational knowledge represents a significant frontier for organizational research with immense and exciting possibilities*” (Yayavaram and Ahuja, 2008, p. 358). In particular, a better understanding of firms’ knowledge search practices helps answer the question of what knowledge search practices may best foster the probability of designing an invention with high technological novelty.

We acknowledge that Nakamura et al. (2015) developed a formal model of combinational process based on combining knowledge *breadth* and *depth* (Laursen and Salter, 2006) and based on IPC analysis. Nevertheless, they did not rely on firm’s knowledge base structure and dynamics, as their research question was driven by the identification of the most suitable sub-domain for a given technological path. Here, we emphasize that the knowledge space of a given company has a particular structure that should be taken in account to explore firm’s knowledge search (Hatchuel and Weil, 2009, 2003; Yayavaram and Ahuja, 2008; Yayavaram and Chen, 2015). **We thus provide here a formal taxonomy of knowledge search modes** based on both the firm’s knowledge base structure and the combination of knowledge components of various distances, with the support of network theory techniques. In that sense, we also help explore questions identified by Phelps et al. (2012) in their literature review on knowledge networks: “*When, why, and how does the intraorganizational knowledge network structure of an organization affect an organizational unit’s knowledge creation and the organization’s knowledge creation?*” (Phelps et al., 2012, p. 137). Then, going a step further, the exploitation/exploration paradigm gives us a chance to measure which knowledge search practices lead to the most novel inventions in terms of technological originality.

Our formal model is computed on the basis of knowledge component couplings and then operationalized through patent data analysis. Taking into account the firm’s knowledge base structure—particularly knowledge component distance and newly absorbed components—and

evaluating its implications for designing inventions necessarily entails a new research design. Firms' knowledge base structures were modeled according to techniques in the literature techniques (Yayavaram and Chen, 2015), but we then used advance network theory tools (e.g., geodesic distance, graph components, weighted edges) to design the overall model. The taxonomy was implemented and tested through a case study of the oil & gas industry.

## 4. TAXONOMY OF INNOVATION SEARCH PRACTICES: FORMAL MODEL AND IMPLEMENTATION THROUGH PATENT ANALYSIS

### 4.1. Firm's knowledge base definition

In order to define a firm's knowledge base, we drew on the common assumption that to design a given invention, firms need to combine knowledge components (e.g., Katila, 2002; Nelson and Winter, 1982). Hence, to design invention  $i$ , a given firm needs to combine  $n$ -tuples of knowledge components (referring to the sequence of all knowledge components that have to be mobilized to produce the invention). We adopted a pairwise logic, following the methodologies of Strumsky and Lobo (2015), Verhoeven et al. (2016), Yayavaram and Ahuja (2008), and Yayavaram and Chen (2015).<sup>3</sup>

Knowledge component couplings can be reinterpreted through (undirected) network theory (Yayavaram and Ahuja, 2008; Yayavaram and Chen, 2015). We acknowledge that a wide stream of literature has used network theory to explain knowledge dynamics (cf. Phelps et al., 2012, for a review) but those were mainly focusing on either inventor level or invention citations level (e.g., Grigoriou and Rothaermel, 2017); here, instead, we followed a similar approach to Yayavaram and Ahuja (2008) and Yayavaram and Chen (2015), who used knowledge component networks for firm-level analysis. Graph  $G_i$  for invention  $i$  is comprised of vertex  $V(G_i)$ , representing the **knowledge components** that have to be mobilized to design the invention and edges  $E(G_i)$ , representing the **knowledge couplings** of the  $n$ -tuple = components (i.e., structure). At a given time  $t$ , the firm may have developed multiple inventions; we are then able to map the **knowledge base** generated by all those inventions at given time by simply unifying the different graphs of each invention. We made the assumption that, following this operation, the weight of each existing vertex and edge can be considered equal to 1.<sup>4</sup>

Over time, a given firm produces diverse inventions by combining previously existing components from time  $t$  and newly accessed knowledge components at time  $t + 1$ .<sup>5</sup> Figure 2 represents an example of a given firm's knowledge base at time  $t+1$ , in which the firms could have designed, for example in time  $t$ , the following inventions: {A,B}, {B;C;D}, {D;E}, {D;F}.

<sup>3</sup> The number of knowledge component pairs (KPs) for invention  $i$  is then:  $KP_i = C_{n_i}^2 = \frac{n_i!}{2(n_i-2)!}$

<sup>4</sup> There is no escalation for knowledge accumulation, as we only focused on knowledge base structure and components.

<sup>5</sup> New vertex and combinations in time  $t+1$  are then:  $NEW\ EDGE_{t+1} = E(G_{t+1}) - (E(G_{t+1}) \cap E(G_t))$  and  $NEW\ VERTEX_{t+1} = V(G_{t+1}) - (V(G_{t+1}) \cap V(G_t))$

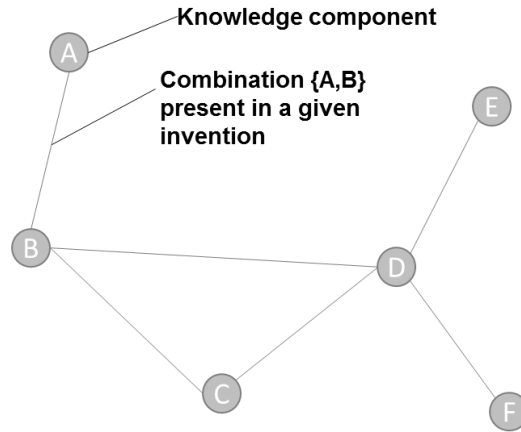


Figure 2 - Knowledge base example following inventions {A,B}, {B,C,D}, {D,E}, and {D,F}

## 4.2. Taxonomy of knowledge search modes

To help in the understanding of knowledge search practices, we developed a taxonomy based on backward reasoning: what would be the effects of a newly designed invention at time  $t+1$  on the previous knowledge base at time  $t$ . The taxonomy was designed based on three key dimensions: (1) the integration, or not, of new knowledge components in the firm's knowledge base; (2) the creation, or not, of new combinations of pre-existing knowledge components; and (3) the distance between the knowledge components in the firm's knowledge base.

- **Refinement mode**

Inventions relying on the **refinement mode** are entirely based on the firm's previous efforts to acquire and combine knowledge. Hence, refinement inventions do not add any new vertex or edge to the previous knowledge structure, as those inventions are only reemploying existing knowledge and combination of knowledge. This definition is in line with Strumsky and Lobo (2015). Inventions designed through this mode are related to local search and can be viewed as a proxy for *knowledge depth*: the firm is incrementally improving its knowledge in an already known and mastered technical discipline or area of expertise. Formally, for a new combinatorial invention  $I_{t+1} = \{1, \dots, n\}$  designed at time  $t + 1$ , with  $d_t^{i-j}$  the shortest path at time  $t$  between components  $i$  and  $j$ :

$$V(G_{t+1}) - V(G_t) = 0$$

$$E(G_{t+1}) - E(G_t) = 0$$

$$\max(d_t^{i-j}) = 1, \forall i, j \in [1, n]$$

- **Clustering mode**

Inventions relying on **clustering mode** are based on combining knowledge components searched locally, already present in the firm's knowledge base, but not yet combined by the firm. Hence, the design leads to at least one new combination of already "close" existing knowledge components. To operationalize the knowledge structure, we computed geodesic distance for each pair of vertexes.<sup>6</sup> We consider that the geodesic distance between two

<sup>6</sup> The path with the minimum number of edges (Kolaczyk, 2009).

knowledge components constitutes an indicator of what can be considered local throughout the firm's knowledge base. For instance, combining knowledge components that have a geodesic distance strictly equal to 2 can be considered local from a firm's inventor's standpoint; the combination is then creating a knowledge cluster.<sup>7</sup> Grigoriou and Rothaermel (2017) followed a similar approach by using average lengths between nodes to proxy the capacity of an inventor to access and recombine existing knowledge. We made the assumption that those elements could describe new product development projects involving, for example, two divisions of a given business unit; this would also guarantee higher modularity for future projects. Formally, for a new combinatorial invention  $I_{t+1} = \{1, \dots, n\}$  designed at time  $t + 1$ , with  $d_t^{i-j}$  the shortest path at time  $t$  between components  $i$  and  $j$ :

$$V(G_{t+1}) - V(G_t) = 0$$

$$E(G_{t+1}) - E(G_t) \geq 1$$

$$d_t^{i-j} = 2, \exists i, j \in [1, n]$$

$$\max(d_t^{k-l}) \leq 2, \forall k, l \in [1, n]$$

- **Absorption mode**

Inventions based on **the absorption mode** are those that draw on new knowledge components that were not part of the knowledge structure at time  $t-1$  and that do not imply new combinations of previously unconnected existing components in the firm's knowledge structure, except for local components (i.e., geodesic distance strictly equal to 2). Hence, those inventions can either be based on a combination of new knowledge components that are completely unconnected to the main knowledge structure, which leads to creation of a new component in the graph (case 1), or they can be based on new knowledge components that are combined with only one existing knowledge component (case 2). This mode refers to the situation in which the firm is sourcing knowledge outside of its knowledge base to complete the design of the invention; most probably, those new distant knowledge components are the results of an explorative knowledge search practice with external partners or of completely new greenfield research projects. Formally, for a new combinatorial invention  $I_{t+1} = \{[1, \dots, m], [n, \dots, o]\}$  designed at time  $t + 1$ , with  $d_t^{i-j}$  the shortest path at time  $t$  between components  $i$  and  $j$ :

$$V(G_{t+1}) - V(G_t) \geq 1$$

$$E(G_{t+1}) - E(G_t) \geq 1$$

Case 1: absorption by additional knowledge component:

$$V(I_{t+1}) \cap V(G_t) = \emptyset$$

Case 2: absorption by connection with existing knowledge structure:

$$d_t^{i-j} \begin{cases} = \emptyset, \exists i, j \in [1, \dots, m] \\ \leq 2, \forall i, j \in [n, \dots, o] \end{cases}$$

---

<sup>7</sup> It could also be viewed as a circuit from a matroid perspective.

- **Recomposition mode**

Inventions based on the **recomposition mode** comprise new combinations of existing knowledge that are not considered local in the firm's knowledge base (i.e., that entail major changes in the structure). Potentially, this category also includes combinations involving new knowledge components. This practice is particularly challenging, as the firm needs to combine knowledge components that are already mastered within the firm but are not considered local from the inventor's point of view. This can occur because of fixation effects (Agogu  and Le Masson, 2014) or lack of communication between silos, for example. This redefinition of the links between familiar and mastered knowledge components can be generated through access to new knowledge components. Formally, for a new combinatorial invention  $I_{t+1} = \{1, \dots, n\}$  designed at time  $t + 1$ , with  $d_t^{i-j}$ , the shortest path at time  $t$  between components  $i$  and  $j$ :

$$V(G_{t+1}) - V(G_t) \geq 0$$

$$E(G_{t+1}) - E(G_t) \geq 1$$

$$d_t^{i-j} \in [3; +\infty[, \exists i, j \in [1, n]$$

Our taxonomy is based on four knowledge search practices. Table 1 provides a synthesis of these four practices with stylized examples.

	Initialization	Refinement mode	Clustering mode	Absorption mode	Recomposition mode
<b>Invention</b>	NA	$I_{t+1} = \{1, \dots, n\}$	$I_{t+1} = \{1, \dots, n\}$	$I_{t+1} = \{[1, \dots, m], [n, \dots, o]\}$	$I_{t+1} = \{1, \dots, n\}$
<b>Period</b>	$t$	$t+1$	$t+1$	$t+1$	$t+1$
<b>Vertex properties</b>	NA	$V(G_{t+1}) - V(G_t) = 0$	$V(G_{t+1}) - V(G_t) = 0$	$V(G_{t+1}) - V(G_t) \geq 1$	$V(G_{t+1}) - V(G_t) \geq 0$
<b>Edge properties</b>	NA	$E(G_{t+1}) - E(G_t) = 0$	$E(G_{t+1}) - E(G_t) \geq 1$	$E(G_{t+1}) - E(G_t) \geq 1$	$(G_{t+1}) - E(G_t) \geq 1$
<b>Geodesic distance properties</b>	NA	$\max(d_t^{i-j}) = 1, \forall i, j \in [1, n]$	$d_t^{i-j} = 2, \exists i, j \in [1, n]$ $\max(d_t^{k-l}) \leq 2, \forall k, l \in [1, n]$	Case 1: $V(I_{t+1}) \cap V(G_t) = \emptyset$ Case 2: $d_t^{i-j} = \begin{cases} \emptyset, \exists i, j \in [1, \dots, m] \\ \leq 2, \forall i, j \in [n, \dots, o] \end{cases}$	$d_t^{i-j} \in [3; +\infty[$ $\exists i, j \in [1, n]$
<b>Stylized example</b>		 $I_t = \{B, C, D\}$	 $I_t = \{A, C\}$	 $I_t = \{A, J\}$	 $I_t = \{A, K, E\}$

Table 1 - Knowledge search modes taxonomy

### 4.3. Implementation of the taxonomy through patent analysis

In this section we detail the implementation of the knowledge search modes taxonomy developed above, using patent data and network theory techniques. We used Patent Classification Codes to compute knowledge base components for a given firm at time  $t$  and applied co-classification data to build the network (i.e., the structure of the knowledge base).

We relied on the International Patent Classification (IPC) codes from the World Intellectual Property Organization, which are more suitable for assessing the recombinant nature of invention than the United States Patent Office Classification (Gruber et al., 2013), as the IPC system uses complete technological information contained in a given patent application while the USPOC system uses only technologies embedded in the claims. Each IPC code is a hierarchical symbol denoting section, class, subclass, main group, and subgroup (WIPO, 2019). The IPC list was first published in 1971; we referred to the 2019 version, which contains 131 classes and 645 subclasses. It has to be highlighted that a patent can be classified in more than one subclass; indeed, for example, in the 2000–2014, period only 15% of USPTO-issued patents can be described as having only one technological functional (Strumsky and Lobo, 2015).

To design the firm's knowledge base, we relied on the IPC subclasses (i.e., IPC-4) as we considered that we only need to account for consequential knowledge components and combinations. This assumption is consistent with practices in the field (e.g., Verhoeven et al., 2016). To model the base, we looked at all the patent applications of a given firm at time  $t$ , due to issues related to international extensions and other patenting specificities, we do not account for patents individually but for families of patents, to analyze inventions per se (Martínez, 2011). We used all patent applications, including for patents that were not granted, under the assumption that if the firm is willing to apply for a patent, due to the costs of application procedures, the knowledge components involved in the given invention have been mastered by the company. We interpret it that if a given patent is not issued, it is due to competitors' previous applications, procedural issues, or incomplete review of the previous state of the art, none of which interferes with what knowledge components are mastered or not by the firm.

In this situation, for a given invention  $I$  at time  $t$ , the invention graph would be represented as follows:

- Vertex  $V(G_t^I)$  : each IPC code of the given invention family;
- Edge  $(G_t^I)$  : each combination of IPC codes of the given invention family (i.e., co-occurrence).

Then, the complete knowledge structure for a given firm is calculated through the unification of all patent families with edge simplification.

To compute the final version of the model, we had to specifically deal with both (1) time constraints, to implement the network dynamic, and (2) mono-class patents. Regarding the former, we needed to compute whether each vertex at time  $t+1$  was new or existed in the previous knowledge base at time  $t$ . To avoid a binary model and to take into account time-lag due to investment and research duration required to master given new knowledge, a new vertex would be considered new for a period of 3 years (e.g., Yayavaram and Ahuja, 2008). On the other hand, it has been shown that knowledge comprising of filed patents is depreciated over time (Park et al., 2006) and that the depreciation rate is industry dependent. We defined a depreciation cycle  $d_f$  that differed depending on studied industry  $f$ ; a given knowledge component's or combination's lifetime in industry  $f$  is equal to  $d_f$ . The lifetime of all knowledge



components or combination of different industries are depreciated at different rates. Vertices and edges were deleted from the knowledge base if there was no other patent family filed during the period with a similar IPC code or combination of IPC codes. Others have also referred to a knowledge component lifetime to analyze dynamic knowledge absorption or usage by firms or inventors (e.g., Grigoriou and Rothaermel, 2017; Yayavaram and Chen, 2015).

As for mono-class patents, some inventions only comprise a unique IPC code; studies focusing on combinatorial innovations do not take in account those patents (e.g., Verhoeven et al., 2016). Our approach is different, as in the initial phase mono-class patents are taken into account for knowledge base computation, knowledge component newness, and lifetime duration calculation. Nevertheless, mono-class patent families are not consistent with our approach to take into account technological originality, and thus are excluded from the final studied sample after knowledge base computation.

## 5. METHODOLOGY

### 5.1. Industrial sector specifics: Oil & gas

We conducted an exploratory study on the energy sector, in particular the oil & gas subsector. The oil & gas industry is key to almost all modern economies because of its major dependency on fossil fuel, significance in the global energy mix and impacts on societies (Korotayev et al., 2018). In this study we are focusing on oil & gas industry innovation practices. In this subsector, R&D efforts are dynamically increasing due largely to three major shifts in the market: (1) the decreasing stock of oil & gas resources, requiring the development of new technologies to find and produce hydrocarbons as they become more difficult to source and produce; (2) major disasters such as Exxon Valdez (1989), Brent Spar (1995), or Deepwater Horizon (2010), which have led to increasing R&D efforts to sustain human and environmental safety; and finally (3) the diversification of major players towards more renewable energy alternatives (cf. Perrons, 2014 for a detailed review of R&D trends in the sector).

Regarding our research questions, on the one hand, firms in the oil & gas industry have shown a dynamic innovation trend over the last decades, while on the other hand, some actors are old, very well-established companies we can assume have accumulated a large knowledge base over time. The subsector thus constitutes a good candidate for research on knowledge search and the design of new inventions.

### 5.2. Data sample

We only focused on major firms in the worldwide market, as we wanted to select firms with a large knowledge base. Firms were selected using the Thomson Reuters Top 100 Energy Report (Thomson Reuters, 2017) subsection on the top 25 companies for the oil & gas subsector. This sample includes five most prominent oil- & gas-integrated companies: companies operating in upstream (exploration and exploitation), midstream (transportation, storage, and processing) and downstream (refining, purifying and marketing and commercial distribution of various products, such as natural gas, kerosene, asphalt and other petrochemicals materials)<sup>8</sup>. We retrieved patent data using the *Clarivate Derwent* database, which includes patents filed in major patent offices (EP, WO, and US); we used the “Optimized

---

<sup>8</sup> The five most prominent industries include: Exxon Mobil, Royal Dutch Shell, Chevron, BP and Total Group

Assignee” function to retrieve patents for the top 25 oil & gas companies. *Clarivate Derwent* conducted in-depth analyses for the 21,000 top worldwide companies in terms of patents, to retrieve filed patents, identify major subsidiaries, and correct company name spelling issues. We were able to retrieve 16 companies’ patent portfolios from among the 25 selected in the sample.<sup>9</sup> With the help of experts from the Technological Intelligence Unit and the Global R&D Patent Unit of one of the major Oil & Gas companies present in the sample, we were able to obtain more fine-grained data due to corrections on subsidiaries; non-completely integrated subsidiaries were excluded from the sample. Patent data were initially retrieved for applications between January 1, 1980 and September 9, 2019.

Ultimately, our sample comprises 16 firms accounting for 125,465 unique patents<sup>10</sup> and 48,662 International Patent Documentation (INPADOC) families (i.e., inventions). Breakdown per company, invention, and patent is available in Appendix 1.

**5.3. Implementation of the taxonomy on the data set**

Using the detailed data sample, we implemented our model to define knowledge search modes. We followed the steps described in Figure 3 and detailed below. The model was developed using R code (through an R Studio Interface) developed by the authors specifically for this study.

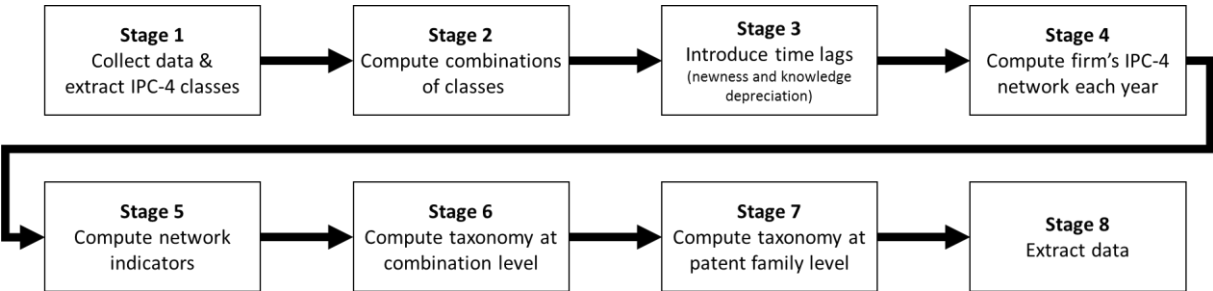


Figure 3 - Stages of model implementation

- **Stage 1: Collect data & extract IPC-4 classes**

IPC-4 classes that were part of a given patent were aggregated at INPADOC patent family level; we computed IPC-4 data for the 48,662 INPADOC patent families. The date used for the INPADOC patent family is the earliest date for any of the patents in the family. Each INPADOC patent family including application year and IPC-4 n-tuples, was assigned to a company portfolio.

- **Stage 2: Compute combination of classes**

<sup>9</sup> Excluded companies included Bharat Petroleum, Gazprom, Hess Corporation, MOL, Lukoil, PKN Orlen, PTT Public Company Limited, Thai Oil, and S-Oil, as for these companies *Clarivate Derwent* did not perform in-depth analysis (i.e., did not have optimized assignee functionality); therefore, we considered that patent data for those companies were not reliable enough to be included.

<sup>10</sup> The total number of patents is 125,465, but the breakdown per company is higher, as some are counted multiple times because of partnerships between companies in the sample.

Based on the list of IPC-4 n-tuples of each INPADOC patent family, a list of IPC-4 pairwise combinations was built. A total of 83,970 combinations and INPADOC patent family pairs were retrieved from the sample.

- **Stage 3: Introduce time lags**

Time lags are introduced to take into account newness of knowledge components and knowledge stock depreciation. Each new IPC-4 or IPC-4 pairwise combination in a firm's patent portfolio is considered new to the firm for a duration of 3 years (cf. section 4.3). Park et al. (2006) analyzed knowledge depreciation over different industries; according to their computation, the depreciation rate for the chemical industry is 13.10% on average, while that for non-metallic minerals is 12.84% in average (Park et al., 2006, p. 125)<sup>11</sup>. We set a knowledge components and combinations lifetime of 8 years;<sup>12</sup> if a given IPC-4 or pairwise combination of IPC-4s was not filed again through another INPADOC patent family in that time, it was deleted from the firm's knowledge base after 8 years. Following the application of time lags, 204,239 combinations of IPC codes were computed. It has to be noted that for a given company, if an IPC-4 is deleted after 8 years due to no further applications but is then filed again after 8 years, it is treated as a new one.

- **Stage 4: Compute firm's IPC-4 network each year**

Based on IPC-4s extracted from mono-class INPADOC patent families and IPC-4 pairwise combinations extracted from multi-class INPADOC patent families, with the application of time lags, IPC-4 networks were computed to design firms' knowledge base. A network graph was generated each year from 1980 to 2019 for each firm in the sample. Because of variations regarding the first year of patenting due on different dates of firm establishment, 558 graphs were generated. An example is given for the company Total S.A. in 1995 in Figure 4 below.<sup>13</sup>

---

<sup>11</sup> The oil & gas industry does not have a dedicated sector in Park et al. (2006). Because of a high reliance of this industry on the chemical industry for downstream activities and on nonmetallic minerals for upstream activities, we relied to the latter.

<sup>12</sup> We computed the average between 13.10 % and 12.84% and computed the time for complete depreciation of a given knowledge component, yielding 8 years.

<sup>13</sup> IPC-4 short descriptions are sourced from Alstott et al. (2017a).

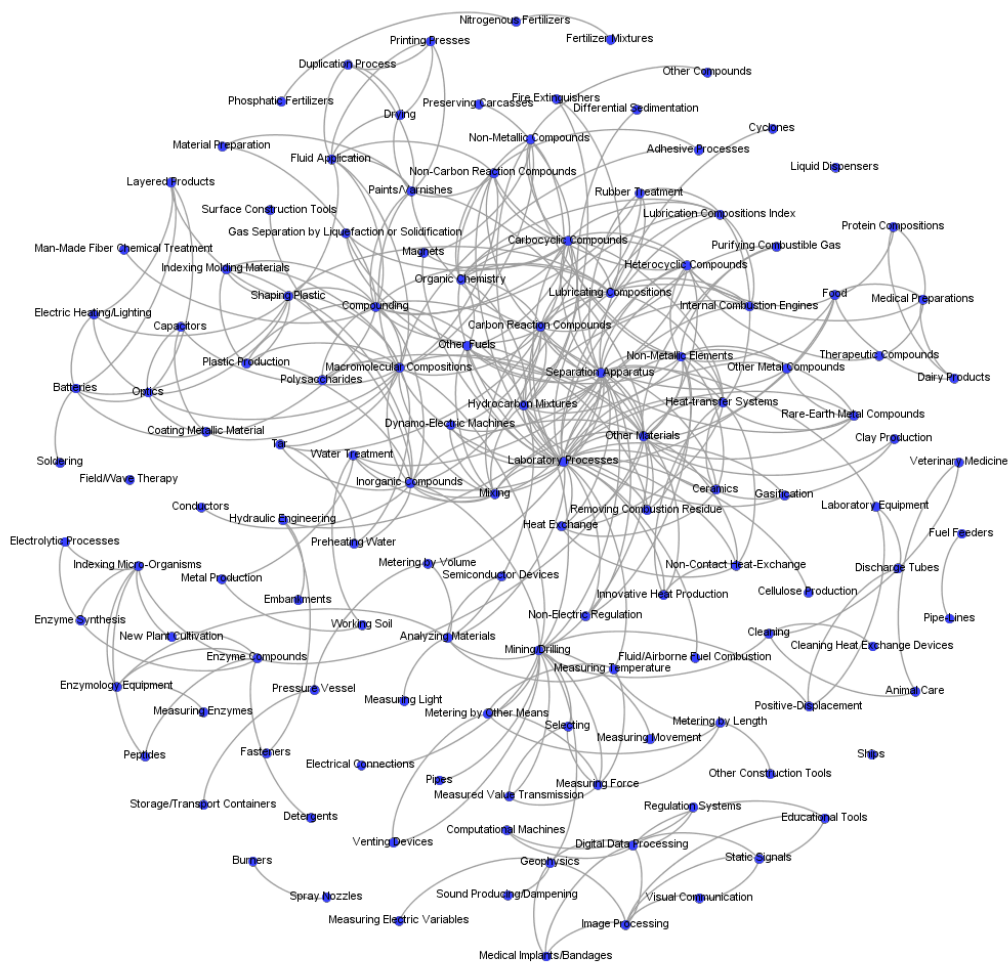


Figure 4 - Knowledge base TOTAL 1995 based on IPC-4 definitions

- **Stage 5: Compute network indicators**

For each network graph, the geodesic distance between any pair of IPC-4s is computed, generating 4,777,396 observations. Furthermore, each vertex of each network graph was classified as belonging to the principal component or to a non-principal component.

- **Stage 6: Compute taxonomy at IPC-4 pairwise combination level**

For each pairwise IPC-4 combination, a first level of analysis is computed to generate the complete taxonomy at INPADOC patent family level at the next stage. Hence, we distinguished between:

- IPC-4 pairwise combinations with only non-new IPC-4s that had been filed for more than 3 years—*refinement pairwise combinations*.
- IPC-4 pairwise combinations with only non-new IPC-4s and with a previous geodesic distance strictly equal to 2—*knowledge cluster enhancement pairwise combinations*.
- New combinations of IPC-4s not related to the principal component—*strict external absorption pairwise combinations*.

- New IPC-4 pairwise combinations with a geodesic distance strictly superior to 2, with interactions only with principal components and including both a new and a non-new IPC-4—*non-strict external absorption*.
- New IPC-4 pairwise combinations with a geodesic distance strictly superior to 2, with interactions only with principal components and including two non-new IPC-4s—*recomposition pairwise combination*.

Furthermore, because of newness lag, the first level of the taxonomy was applied for a duration of 3 years, meaning that, for example, if the combination  $\{A, B\}$  was designated in the year 1980 as an *absorption* combination, this was also the case for years 1981 and 1982.

- **Stage 7: Compute taxonomy at patent family level**

As INPADOC patent families may comprise multiple combinations depending on the size of the IPC-4 n-tuples, we defined an algorithm to apply the complete taxonomy. The steps are applied successively to the stock of IPC-4 pairwise combinations per INPADOC patent family.

- Step 1: At least one *strict external absorption pairwise combination* → **absorption INPADOC patent family**;
- Step 2: At least one *recomposition pairwise combination* (entailing no *absorption pairwise combination*) → **recomposition INPADOC patent family**;
- Step 3: At least one *non-strict external absorption* pairwise combination (entailing no *strict external absorption pairwise combination* or *recomposition pairwise combination*) → **absorption INPADOC patent family**;
- Step 4: At least one *knowledge cluster enhancement pairwise combination* (entailing no step 1 to 3 combinations) → **knowledge cluster enhancement INPADOC patent family**;
- Step 5: other cases leading to at least one *refinement pairwise combination* → **refinement INPADOC patent family**.

- **Stage 8: Data extraction**

The final data set was extracted to analyze the different knowledge search modes throughout the top 16 companies in our sample of the oil & gas subsector. As we were focusing on knowledge bases that were dynamically changing over time, we needed to account for time to allow network stabilization. We thus used the sample of patent families only after one complete cycle of knowledge depreciation (i.e., 8 years); furthermore, INPADOC patent families filed between 2017 and 2019 were deleted due to the delay between the moment of patent publication and submission to ensure data consistency. After these restrictions, the sample contains 32,594 INPADOC patent families. Finally, in line with previous studies (see section 4.3), we excluded INPADOC patent families with only a single class, as we were focusing on combinations. The final sample contained 19,825 INPADOC patent families

## 5.4. Invention technological originality measurement

The measure of technological invention originality is a complete innovation management field in itself, and it is not our objective here to provide an exhaustive assessment (see Squicciarini et al., 2013 for a review). Generally, when using patent data, scholars are accounting for patent citations in order to proxy the value of the invention, such as its commercial potential, usefulness, or social welfare benefits (Hall et al., 2000; Harhoff et al., 1999; Jaffe and de Rassenfosse, 2010; Trajtenberg, 1990). It has nevertheless been shown

that patent citation is subject to several biases (such as time lag, variations in citations by firm size, auto-citation, etc.); in order to overcome those issues, authors have used semantic analysis techniques to measure the novelty of a given invention (e.g., Arts and Fleming, 2016; Hain et al., 2018). In this study, our focus was on patent classification classes and then the combinatorial originality of combinations; as a result, we looked for a technological originality proxy using patent classification classes. One of the main advantages of this approach is that the measure would not be affected by citations time lag and helping it would be recognized as a valuable invention.

We chose to rely on global Patent Classification Class mapping (Leydesdorff et al., 2014; Schoen et al., 2012). Specifically, we used an indicator derived from Alstott et al. (2017a, 2017b), who developed a methodology to compute technology network structure and draw a measure of proximity between each pairwise technological combination at IPC-4 level. In more detail, the technology network was computed by reviewing 3,911,050 utility patents issued from 1976 to 2010 by the United States Patent and Trademark Office. The measure of technological proximity aggregated multiple proximity indicators: direct citations, co-citation, cosine similarity, co-occurrence and co-classification, along with several normalization measures in order to correct impinging factors. In particular, the authors showed that on one hand, the technology network is sparse, as very few pairs of technology classes are significantly related, while on the other hand, the network is very stable over time. We called this the *Alstott Score*: the proximity between two IPC-4 patent classes, which was normalized between [0; 1].

In our study, we used an *Alstott Derived Originality Indicator* (ADOI). As the technology network is considered stable, we made the assumption that combining two IPC-4 classes that were not close in Alstott et al.'s network was an adequate indicator of technological originality. Thus, the ADOI for the pairwise combination of IPC-4s  $i$  and  $j$  was computed as follows:

$$ADOI^{i-j} = 1 - Alstott\ Score^{i-j}$$

As an INPADOC patent family can comprise several pairwise combinations, we used an arithmetic mean to compute the ADOI score. The Alstott Score database was sourced from a file made publicly available by those authors following their publication.<sup>14</sup>

Our sample contained 496 single IPC-4s, of which 9 had not been computed by Alstott et al. (2017a). A total of 48 INPADOC patent families were impacted by those 9 classes and were excluded from our sample. The final sample thus comprised 61,901 IPC-4 pairwise combinations, of which 7,261 were unique IPC-4 pairwise combinations.

## 6. RESULTS

### 6.1. Firm's knowledge base features

The final sample comprised 19,863 INPADOC patent families representing 50,776 patent applications filed between 1989 and 2016. We reviewed the list of IPC-4 classes present in the sample's applications and found it consistent with the oil & gas industry (Appendix 2). It has to be noted that the number of INPADOC patent family applications in the sample varied over time (Appendix 3); this is presumably because firms employ different patenting practices and some started to patent only after 1989. We noted an increase in patent family applications

---

<sup>14</sup> <https://doi.org/10.5281/zenodo.1035448>

since 1998, in line with Perrons' (2014) work on innovation effort acceleration in an oil & gas firm R&D.

The 16 firms in our sample also recorded variations in terms of the number of applications filed per year and of knowledge base features (Appendix 1). The total number of INPADOC patent family applications varies from 5,596 for Exxon to 65 for Suncor Energy. Exxon Mobil is a highly established company formerly called the Standard Oil Company, founded in 1870,<sup>15</sup> and is highly integrated in the oil & gas supply chain (upstream, midstream, and downstream). Exxon Mobil employed more than 70,000 people worldwide in 2017. On the opposite end, Suncor Energy, a Canadian company established in 1917, operates almost exclusively upstream and employed 12,480 people in 2019.<sup>16</sup> Those differences in the firm size and financial capacity impact patenting practices and subsequently the firm's knowledge base features, such as potential number of knowledge components and maximum number of potential combinations. Variations can also be explained by a firm's intellectual property unit's internal practices and can be subject to major variations over the years. Hence, the average size of the knowledge base and its structure varies between the different companies (Table 2; see also Appendix 4 & Appendix 5 for more details).

	Average no. patent families filed per year	<i>Standard deviation</i>	Average no. IPCs in the knowledge base	<i>Standard deviation</i>	Average no. IPCs pairwise combinations in the knowledge base	<i>Standard deviation</i>	Total number of patent families in the sample
Exxon Mobil	199.9	47.1	235.3	26.2	1,174.4	191.5	5,596
Royal Dutch Shell	139.0	31.5	228.2	30.2	982.6	261.1	3,891
Chevron	74.5	32.0	166.4	24.5	580.7	132.8	2,087
BP	58.9	32.3	185.1	51.2	598.0	224.2	1,648
Saudi Basic Industries Corp.	65.5	78.2	75.0	74.0	348.0	417.4	1,638
Total Group	57.5	28.9	129.2	43.3	435.2	244.8	1,609
ConocoPhillips	40.6	24.7	139.1	38.1	387.7	132.3	1,138
Eni	36.3	14.3	102.7	34.3	323.1	146.2	1,017
Equinor	14.1	6.7	53.7	30.2	95.9	69.7	380
SK Innovation Corp.	41.0	13.3	69.4	42.5	255.7	161.5	246
Reliance Industries	9.5	11.7	28.2	34.7	69.0	101.4	181
PetroChina	9.0	6.2	27.9	19.3	35.9	33.3	117
Repsol	4.9	2.9	33.1	14.5	49.9	25.7	88
Indian Oil Corp.	5.7	5.5	21.3	16.1	32.9	31.9	86
INPEX Corp.	6.3	6.6	12.2	9.6	30.6	28.6	76
Suncor Energy	3.1	2.0	22.0	8.0	42.7	34.6	65
<b>Average / Total</b>	<b>47.9</b>	<b>54.4</b>	<b>95.5</b>	<b>76.2</b>	<b>340.1</b>	<b>350.3</b>	<b>19,863.0</b>

Table 2 – Firm's knowledge base features breakdown

## 6.2. Knowledge search practices

As per our taxonomy, we estimated the proportions of the respective knowledge search mode in the sample: (1) refinement, (2) clustering, (3) absorption, and (4) recomposition. Results are presented in Table 3. The refinement knowledge search mode was the most used (47% of the sample), consistent with the results of our expert interviews. This indicates that firms in this industry are mainly developing in-depth knowledge in or applying existing scientific and technological disciplines in order to incrementally develop new inventions. The second most common knowledge search mode was the absorption mode (38% of the sample). This

<sup>15</sup> Company's corporate website.

<sup>16</sup> Company's corporate website.

was also consistent with our expert interviews, reflecting the promotion of open innovation projects across the industry. As noted, due to major changes in the field and the increase in firm innovation efforts, companies are looking more to greenfield projects; impacts on invention may be driven, for example, by new start-up acquisitions that became fully integrated in the organization, bringing non-traditional research areas for the company. Finally, clustering and recomposition modes were the least used modes, again consistent with the expert interviews. Those knowledge search modes face particular silo issues because they require new combinations of the mastered knowledge components in the firm’s knowledge base, which probably indicates that the knowledge components are mastered by a different team or business unit and not by the one currently handling the invention.

	No. patent families	Repartition
Refinement mode	9,336	47.0%
Absorption mode	7,534	37.9%
Clustering mode	2,304	11.6%
Recomposition mode	689	3.5%
<b>Total</b>	<b>19,863</b>	<b>100%</b>

Table 3 - Knowledge search modes in the sample

We also looked at the breakdown of knowledge modes by firm and at representations of each mode in terms of patent family application portfolios. As a result, we were able to highlight huge variations in knowledge search practices across the firms in the sample, as shown in Figure 5. As, there was variation in the number of patents filed each year per firm, we also detailed total number of patent families filed through each knowledge search mode and total number of associated INPADOC patent family applications in Table 4.

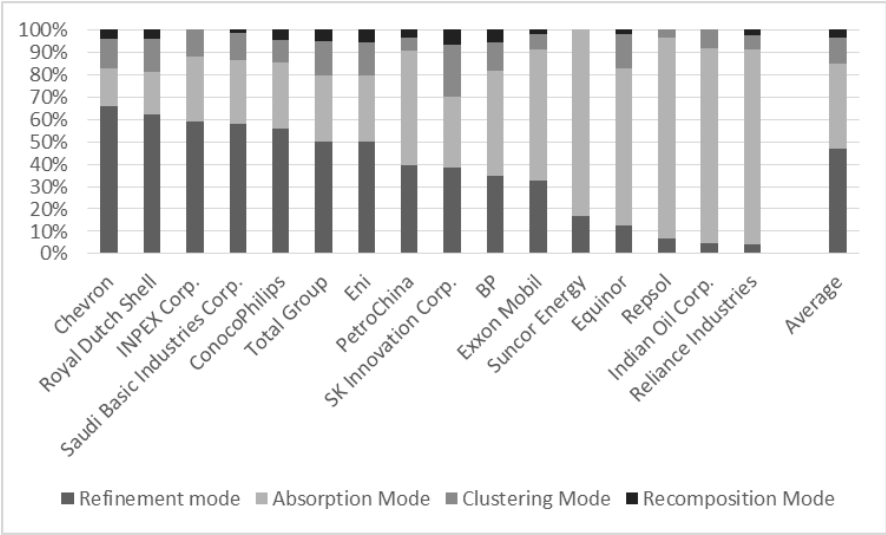


Figure 5 - Firm's INPADOC patent family applications per Knowledge Search Mode



	Refinement mode		Clustering mode		Absorption mode		Recomposition mode		Total no. patent families
	No. patent families	% of firm's portfolio	No. patent families	% of firm's portfolio	No. patent families	% of firm's portfolio	No. patent families	% of firm's portfolio	
Exxon Mobil	1,818	32.5%	387	6.9%	3,295	58.9%	96	1.7%	5,596
Royal Dutch Shell	2,422	62.2%	587	15.1%	726	18.7%	156	4.0%	3,891
Chevron	1,371	65.7%	270	12.9%	359	17.2%	87	4.2%	2,087
BP	570	34.6%	206	12.5%	779	47.3%	93	5.6%	1,648
Saudi Basic Industries Corp.	945	57.7%	196	12.0%	470	28.7%	27	1.6%	1,638
Total Group	804	50.0%	245	15.2%	476	29.6%	84	5.2%	1,609
ConocoPhillips	636	55.9%	114	10.0%	335	29.4%	53	4.7%	1,138
Eni	508	50.0%	148	14.6%	301	29.6%	60	5.9%	1,017
Equinor	48	12.6%	57	15.0%	267	70.3%	8	2.1%	380
SK Innovation Corp.	94	38.2%	57	23.2%	79	32.1%	16	6.5%	246
Reliance Industries	8	4.4%	11	6.1%	157	86.7%	5	2.8%	181
PetroChina	46	39.3%	7	6.0%	60	51.3%	4	3.4%	117
Repsol	6	6.8%	3	3.4%	79	89.8%	-	0.0%	88
Indian Oil Corp.	4	4.7%	7	8.1%	75	87.2%	-	0.0%	86
INPEX Corp.	45	59.2%	9	11.8%	22	28.9%	-	0.0%	76
Suncor Energy	11	16.9%	-	0.0%	54	83.1%	-	0.0%	65
<b>Total</b>	<b>9,336</b>	<b>47.0%</b>	<b>2,304</b>	<b>11.6%</b>	<b>7,534</b>	<b>37.9%</b>	<b>689</b>	<b>3.5%</b>	<b>19,863</b>

Table 4 - Repartition of the knowledge search modes in the sample (thousands of INPADOC patent families)

### 6.3. Invention technological originality & knowledge base innovative potential

We computed the ADOI score for each patent family in the sample. There were only a few patent families with high ADOI scores: 2.1% had an ADOI score of higher than 0.9, which indicates high technological originality (Appendix 6). Very low scores were common: 59.3% of the patent families had an ADOI score below 0.2 (Figure 6). This is consistent with the findings of Alstott et al. (2017a), which show using a large set of patents that only a few classes are connected together.

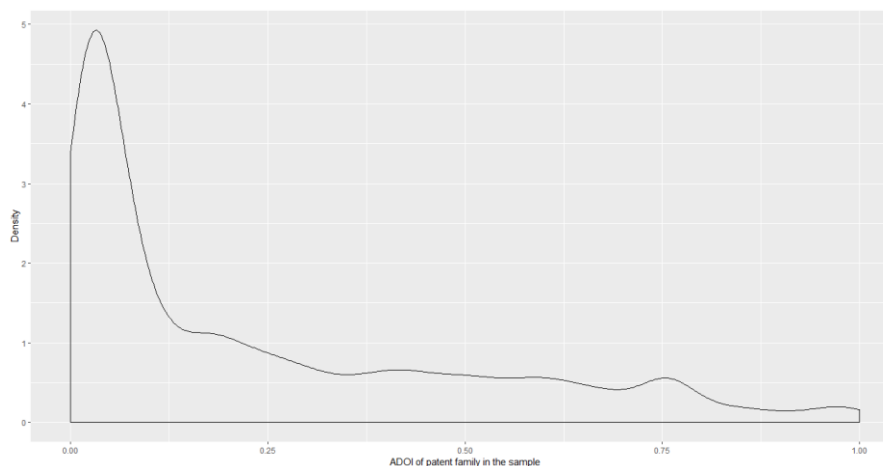


Figure 6 - ADOI density in the sample

Furthermore, it has to be noted that firms recorded high variations in terms of the overall ADOI average score based on patent application families filed each year (Table 5). Suncor Energy and Equinor led the portfolio ADOI score ranking, while Saudi Basic Industries Corp. and INPEX Corp had the lowest scores. The traditional top 5 companies were led by Chevron

and Total Group, while Exxon Mobil scored more poorly. However, it has to be highlighted that standard deviation was very high for this measure and that there were substantial yearly variations for some firms, such as Suncor Energy, Equinor, and Indian Oil Corp. (Appendix 7).

	ADOI of firm's filed patent families	95% confidence interval	
		min.	max.
Suncor Energy	0.391	0.324	0.459
Equinor	0.299	0.272	0.326
SK Innovation Corp.	0.286	0.252	0.320
Chevron	0.281	0.269	0.293
Total Group	0.260	0.247	0.274
Royal Dutch Shell	0.258	0.250	0.267
Indian Oil Corp.	0.253	0.184	0.322
BP	0.253	0.240	0.266
Repsol	0.245	0.189	0.301
Exxon Mobil	0.231	0.224	0.238
Eni	0.230	0.214	0.245
Reliance Industries	0.223	0.185	0.261
PetroChina	0.216	0.169	0.263
ConocoPhillips	0.195	0.181	0.209
Saudi Basic Industries Corp.	0.187	0.175	0.199
INPEX Corp.	0.125	0.085	0.165
<b>Average</b>	<b>0.242</b>	<b>0.238</b>	<b>0.246</b>

Table 5 - Average ADOI score for patent family applications—breakdown per company

We also looked at the ADOI potential induced by the firm's knowledge base, which is less sensitive to patenting practices (Table 6) and can be interpreted as showing innovation potential related to the firm's knowledge base originality itself. The ranking is different from the ADOI average based on filed patent family; by the new measure, Exxon Mobil and Royal Dutch Shell were leading the market (more details are available Appendix 8).

	ADOI of firm's knowledge base average	95% confidence interval	
		min.	max.
Exxon Mobil	0.616	0.605	0.627
Royal Dutch Shell	0.609	0.597	0.620
Saudi Basic Industries Corp.	0.599	0.580	0.618
Chevron	0.590	0.575	0.606
BP	0.558	0.543	0.573
Total Group	0.555	0.537	0.573
Indian Oil Corp.	0.545	0.478	0.612
SK Innovation Corp.	0.529	0.496	0.562
ConocoPhillips	0.528	0.510	0.546
Eni	0.512	0.492	0.532
Reliance Industries	0.478	0.438	0.517
Equinor	0.469	0.435	0.503
Suncor Energy	0.468	0.417	0.519
INPEX Corp.	0.446	0.380	0.512
Repsol	0.446	0.391	0.500
PetroChina	0.350	0.286	0.414

Table 6 - Average ADOI score for firm's knowledge base—breakdown by company

#### 6.4. Knowledge search modes and technological originality

In this section we analyzed what knowledge search mode conduced to the highest originality score for the designed inventions. We crossed the two variables *Knowledge search mode*, computed according to our taxonomy, and *ADOI per patent family applications*. The detailed analysis results and t-tests are presented in Table 7 and a synthesis in Figure 7.

	ADOI average score per filed patent families	95% confidence interval		p-value
		min.	max.	
Refinement mode	0.160	0.156	0.165	p < 0.001
Absorption mode	0.279	0.273	0.285	p < 0.001
Clustering mode	0.387	0.376	0.397	p < 0.001
Recomposition mode	0.466	0.448	0.484	p < 0.001

Table 7 - Analysis ADOI score and knowledge search mode

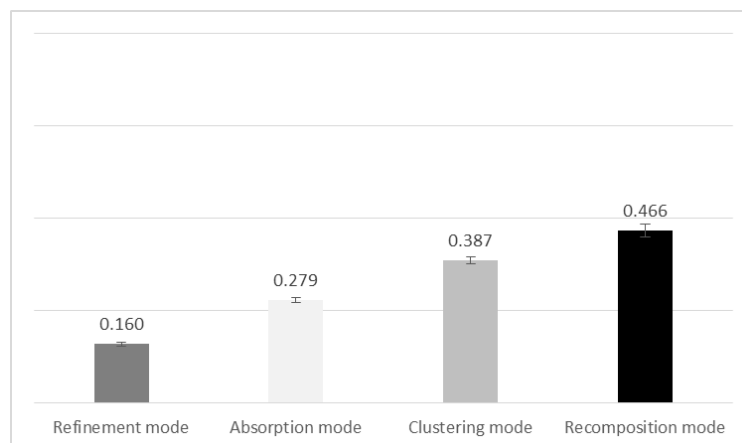


Figure 7 - Graph of ADOI score and knowledge search mode

Thus, for the oil & gas industry, we obtained a clear hierarchy for which knowledge modes gain the highest scores for technological originality. The refinement mode is the least helpful in this regard, with a very low average score of 0.16; at the opposite end, recomposition is the most helpful mode for technological originality in this industry, and is also the mode that accounted for the fewest patent families (3.5% of all the patent families in the sample). Clustering and absorption modes were respectively the second- and third-most helpful knowledge search practices in terms of outcome technological originality.

## 7. DISCUSSION & CONCLUSION

### 7.1. Theoretical implications

In this research, we examined firms' knowledge search practices to support the design of new inventions. We departed from the extant literature by (1) proposing a formal taxonomy of how firms combine knowledge components to design new inventions depending on their degree of reliance on their own knowledge base components and structure and (2) by exploring the relationship between the design of technologically original inventions and the selected knowledge search mode. The formal model was operationalized using patent data and network theory techniques and was applied to a set of companies in the oil & gas industry. The study findings have several important implications for research in the innovation and design field in this industry and in general, as explained next.

First, extant innovation literature shows that firms can combine knowledge components searched locally or distantly in order to design new inventions. Yayavaram and Ahuja (2008) and Yayavaram and Chen (2015) highlighted that firms need to focus both on finding new knowledge components and on reconfiguring the couplings between those components. However, in the literature, no in-depth analysis of how firms rely both on their previous knowledge base—components and structure—and on sourcing knowledge externally to design new inventions was found. We developed a taxonomy based on four knowledge search modes describing knowledge component combinations for a given invention: (1) refinement, (2) clustering, (3) absorption and (4) recomposition. The taxonomy was elaborated through backward reasoning by exploring the effects of a new invention on the firm's previous knowledge base structure. Our model sheds light on both the design of new inventions and the associated creative process in relation to the firm's knowledge base. The taxonomy was implemented by relying on patent data: we used the International Patent Classification (IPC)

to proxy knowledge components and IPC patent co-classifications to proxy knowledge component structures.

Second, we showed to what extent firms relied on the four knowledge search modes in a specific industry, oil & gas. Based on the analysis of 16 key players in this market, we showed that firms mainly designed new inventions by relying solely on their already mastered knowledge components and pre-existing combinations, as the refinement mode accounted for 47.0% of our sample. This gives new insight on Strumsky and Lobo (2015), who found that, across all firms in all industries, 52% of designed inventions reuse a set of similar knowledge components and associated structures existing in the USPTO database. Here, we showed that in addition, firms were reusing sets of similar knowledge components and associated structures existing in their own knowledge base. It has to be highlighted that innovation in the oil & gas industry has been driven by incremental innovation for a long time; it has even been considered an industry of “medium-low R&D intensity” (Moncada-Paternò-Castello et al., 2010, p. 527). When reviewing the technological originality of the refinement knowledge search mode, we found that it scored the lowest of the four tested modes as well as very low in absolute terms (0.16). This means that the set of combined knowledge components used when designing new inventions had been already extensively used in global technological history as sourced through USPTO-filed inventions.

Third, we reviewed the absorption knowledge search mode, which consists in designing new inventions by integrating new knowledge components without major changes in the previous knowledge structure (apart from the new knowledge components themselves). As this implies, this knowledge search mode requires the usage of distant-searched knowledge. Either of two assumptions, or one assumption encompassing two alternatives, can be made regarding the source of this knowledge: either the new knowledge component has been developed through completely new greenfield research projects within the company, or it has been sourced externally by firm acquisitions or partnerships (academia, inter-industry or intra-industry). It is highly probable that the latter approach is more extensively used by firms, as developing new greenfield projects with almost no relations with an established firm’s previous expertise would be a rare event. The absorption mode was widely diffused among the 16 firms in our sample (at 37.9%). It echoes the concept introduced by Cohen and Levinthal (1990) of *absorptive capacity*, which assumes that firms (1) recognize and understand new external knowledge, (2) assimilate valuable external knowledge, and (3) apply assimilated external knowledge (Lane et al., 2006). Among the top 5 oil & gas firms, Exxon Mobil and BP have particular expertise in this domain, as respectively 58.9% and 47.3% of their inventions are designed through this knowledge search mode (making it their most used design mode). Furthermore, as knowledge components in this mode are mainly sourced externally, it refers to open innovation practices (Chesbrough, 2003), practices that firms in the oil & gas industry in the sample have also extensively embraced, as highlighted by our expert interviews and by Perrons (2014). Nevertheless, on average, technological originality score is relatively low for this mode (i.e., 0.279), in particular compared to the recombination mode. This means that, on average, the knowledge components sourced and combined by firms using this mode have also previously been extensively combined in the global technical landscape. Firms that extensively rely on this mode could face difficulties recognizing radically new industry knowledge, as supported by Perrons’ (2014) finding that major firms in this sector prefer to be “fast followers” regarding new technology advancements due to high cost implications in case of failure. However, these firms could also face difficulties in the transformative learning stage when mirroring new knowledge components using more familiar firm components.

Fourth, the knowledge search mode that conduces to the highest technological originality score is the recomposition mode, which implies new inventions based on major changes

regarding firm's knowledge base structure, potentially associated with new knowledge components. This mode was the least represented on average in the sample, as it only accounted for 3.5% of the designed inventions. In fact, 7 firms out of 16 in the sample filed fewer than 10 INPADOC family patent applications through that knowledge search mode. Furthermore, it was consistently the least developed knowledge search mode among the 16 firms in the sample, and SK Innovation Corp., BP, and Total Group were only three firms with more than 5% of inventions designed through it in their portfolios. The low occurrence of this mode in our sample seems to show that designing new inventions through this mode is particularly challenging for firms. Indeed, our expert interviews showed that even in the top 5 companies, historical silos were strong (for example, between oil & gas exploration business units or refining & chemical business units). Mirroring knowledge components that the firm is familiar with but that have to be considered distant from the firm's inventor would be a complex task, but it has been shown in the literature that combining knowledge the firm is familiar with and new knowledge components is generally associated with high innovation performances (Ahuja and Lampert, 2001; Katila and Ahuja, 2002; Yayavaram and Chen, 2015). Our results for inventions designed through the recomposition mode when specifically including new knowledge components, support these findings (Appendix 9). Furthermore, we also showed that inventions designed through the recomposition mode without including new knowledge components also recorded high technological originality scores (0.44 on average, higher than the three other modes—see Appendix 9). While combining knowledge components with which the firm is familiar has in the past been mainly associated with learning traps (Ahuja and Lampert, 2001), focusing on knowledge couplings and the distance between combined knowledge components is a promising new areas of research. The findings also highlight the fact that the distinction between local and distant knowledge at firm level should also be envisioned at firm's inventor level.

## **7.2. Practical implications**

The study helps us understand how firms combine knowledge either from their knowledge base or gained through new knowledge development or external sourcing. The methodology can be leveraged by firms to identify and understand their main modes of knowledge search in order to design new inventions and processes as well as to understand how these modes have evolved over time and how their competitors are designing new inventions in a given industry. It could also provide support to design organizational learning process and set strategic objectives in terms of the knowledge search modes that the firm may undertake in the future. Finally, it can be leveraged as a tool to better understand and predict invention dynamics in the sector.

The study also sheds light on silo issues in organization learning. As recomposition mode conducted to the highest technological originality score, this mode favors organizations that are able to share their expertise towards their silos, without necessarily accessing new external knowledge components. Furthermore, it also implies a challenge to the management of open innovation strategies, as we consider that open innovation is the main model associated with the access to new knowledge components. For example, external collaborations should be driven not only by absorbing external knowledge to develop a given invention or to answer temporary needs; in addition, to maximize technological originality, external collaborations should be managed by gathering multiple firms' expertise with the externally sourced new components. It should also be highlighted that open innovation strategies and organizational learnings could focus more on bridging firm's expertise among silos, potentially by the usage of external knowledge components.

### **7.3. Limitation and further research**

The first limitation of the study concerns the generalizability of the results. The knowledge search taxonomy developed was applied only to one industry and to a restricted sample of firms. Nevertheless, we were able to perform the analysis on a large dataset of inventions (19,863), and we conducted explorative semi-structured expert interviews to validate the model's consistency. Extending our methodology through a larger sample of firms and industries will be a worthwhile avenue for future research. In particular, in-depth qualitative analysis of organizational implications of each knowledge search mode could help the understanding of knowledge search practices.

Second, we only focused here on the technological originality of the studied inventions, whereas other techniques could be applied to measure technological originality and examine whether similar results are found when using, for example, semantic analysis or citation data. In addition, other innovation performance indicators could be reviewed at firm level, in particular financial data or numbers of new launched products. We also see room for further research exploring mitigating variables, such as knowledge complexity, R&D intensity, spending, or reliance on basic science. Further research could also look at how mastering a combination of the four knowledge search modes at different levels could favor R&D efficiency.

Third, the findings contribute to the absorptive capacity and open innovation fields, as the various knowledge search modes encompass new knowledge components or combinations of knowledge components in the firm's knowledge base, some of which may have been sourced externally. We did not control for partnership effects or acquisitions. Looking at the extent to which these practices help the design of original inventions and how this relates to knowledge search modes could also be fruitful areas of research. Furthermore, heeding the call of Yayavaram and Ahuja (2008) to look closely at both the combination of knowledge and knowledge components, it would be interesting to look at whether open innovation practices favor new knowledge component sourcing or new combinations.

Fourth, our model is dependent on the usage of patents as a proxy for firm's knowledge base. Patenting activities, however, are subject to several biases, in particular related to competitive landscape and firm's intellectual property business units' practices. Other research could benefit from the usage of a different proxy for firm's knowledge base, such as academic publications, or from more fine-grained qualitative analysis. It could be promising in particular to explore the case of small firms with no or limited patent portfolios.

Fifth, the knowledge clustering mode received a high ADOI score, an intriguing research result. In particular, knowledge components combined were familiar from the firm's standpoint and the knowledge was considered local also from the firm's inventor's standpoint. The low share of the clustering knowledge mode in the firm's inventions portfolio could indicate that firms are highly fragmented and silos effects are impacting the design of new inventions at a sub-business-unit level. Further research could benefit from exploring this mode in depth to determine what causes this effect.

In summary, this study has yielded several findings that can be useful for both theory and practice. We have argued that an in-depth taxonomy of how firms are relying on their knowledge base and new knowledge acquisition to develop new inventions, focusing both on knowledge components and knowledge structures and their effects, will help in the understanding of the innovation process. We also showed that designing inventions that combine knowledge components that the firm is familiar with but that are considered distant from a firm's inventor's perspective leads to high technological originality. A greater

understanding of the organizational learning practices associated with this finding may offer great potential to understand firms' innovation performance.

## 8. REFERENCES

- M. Agogu , A. Kazakci, A. Hatchuel, P. Le Masson, B. Weil, N. Poirel, M. Cassotti, The impact of type of examples on originality: explaining fixation and stimulation effects, *J. Creat. Behav.* 48 (2014) 112. <https://doi.org/10.1002/jocb.37>
- M. Agogu , P. Le Masson, Rethinking ideation: a cognitive approach of innovation lock-ins, in: *The International Society for Professional Innovation Management, Academy of Management, Vancouver, BC, 2014.* <https://doi.org/https://doi.org/10.5465/ambpp.2015.15303ab> abstract
- G. Ahuja, R. Katila, Technological acquisitions and the innovation performance of acquiring firms: a longitudinal study, *Strateg. Manag. J.* 22 (2001) 197 220. <https://doi.org/10.1002/smj.157>
- G. Ahuja, C.M. Lampert, Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions, *Strateg. Manag. J.* 22 (2001) 521 543. <https://doi.org/10.1002/smj.176>
- J. Alstott, G. Triulzi, B. Yan, J. Luo, Mapping technology space by normalizing patent networks, *Scientometrics* 110 (2017a) 443 479. <https://doi.org/10.1007/s11192-016-2107-y>
- J. Alstott, G. Triulzi, B. Yan, J. Luo, *Inventors' explorations across technology domains*, Cambridge University Press, Cambridge, England, 2017b. <https://doi.org/10.2139/ssrn.2936709>
- C. Andriopoulos, M.W. Lewis, Managing innovation paradoxes: ambidexterity lessons from leading product design companies, *Long Range Plann.* 43 (2010) 104 122. <https://doi.org/10.1016/j.lrp.2009.08.003>
- A. Arora, S. Belenzon, A. Pataconi, The decline of science in corporate R&D, *Strateg. Manag. J.* 39 (2018), 3 32. <https://doi.org/10.1002/smj.2693>
- W.B. Arthur, The structure of invention, *Res. Policy* 36 (2007) 274 287. <https://doi.org/10.1016/j.respol.2006.11.005>
- S. Arts, L. Fleming, Paradise of novelty—or loss of human capital? Exploring New fields and inventive output, *Organ. Sci.* 29 (2016) 1074 1092. <https://doi.org/10.1287/orsc.2018.1216>
- R. Buderl, Engines of tomorrow: how the worlds best companies are using their research labs to win the future, Simon and Schuster Ltd., New York, NY, 2004.
- B. Cabanes, Mod liser l' mergence de l'expertise et sa gouvernance dans les entreprises innovantes : des communaut s aux soci t s proto- pist miques d'experts, Mines Paristech - Paris Sciences & Lettres, Paris, France, 2017. doi:pastel.archives-ouvertes.fr/tel-01743890
- A. Cammarano, F. Michelino, E. Lamberti, M. Caputo, Accumulated stock of knowledge and current search practices: the impact on patent quality, *Technol. Forecast. Soc. Change* 120 (2017) 204 222. <https://doi.org/10.1016/j.techfore.2016.12.019>
- H. Chesbrough, *Open innovation, the new imperative for creating and profiting from technology.* Harvard Business Review Press, Cambridge, MA, 2003.
- H. Choi, J. Shin, W.S. Hwang, Two faces of scientific knowledge in the external technology search process, *Technol. Forecast. Soc. Change* 133 (2018) 41 50. <https://doi.org/10.1016/j.techfore.2018.02.020>
- W.M. Cohen, D.A. Levinthal, Absorptive capacity: a new perspective on learning and innovation, *Adm. Sci. Q.* 35 (1990) 128 152. <https://doi.org/10.2307/2393553>
- W.M. Cohen, D.A. Levinthal, Innovation & learning: the two faces of R&D, *Econ. J.* 99 (1989) 569 596. <https://doi.org/10.1017/CBO9781107415324.004>
- A.C. Cooper, D. Schendel, Strategic responses to technological threats, *Bus. Horiz.* 19 (1976) 61 69. [https://doi.org/10.1016/0007-6813\(76\)90024-0](https://doi.org/10.1016/0007-6813(76)90024-0)
- K.B. Dahlin, D.M. Behrens, When is an invention really radical?: Defining and measuring technological radicalness, *Res. Policy* 34 (2005) 717 737. <https://doi.org/10.1016/j.respol.2005.03.009>
- G. Dosi, Sources, Procedures, and Microeconomic Effects of Innovation, *J. Econ. Lit.* 26 (1988) 112 1171. <https://doi.org/http://www.jstor.org/stable/2726526>
- G. Dosi, Technological paradigms and technological trajectories, *Res. Policy* 11 (1982) 147 162. [https://doi.org/10.1016/0048-7333\(93\)90041-F](https://doi.org/10.1016/0048-7333(93)90041-F)
- L. Fleming, Recombinant uncertainty in technological search, *Manag. Sci.* 47 (2001) 117 132. <https://doi.org/10.1287/mnsc.47.1.117.10671>
- L. Fleming, K. Frenken, The evolution of inventor networks in the Silicon Valley and Boston regions, *Adv. Complex Syst.* 10 (2007) 53 71. <https://doi.org/10.1142/S0219525907000921>
- L. Fleming, O. Sorenson, Science as a map in technological search, *Strateg. Manag. J.* 25 (2004) 909 928. <https://doi.org/10.1002/smj.384>
- L. Fleming, O. Sorenson, Technology as a complex adaptive system: evidence from patent data, *Res. Policy* 30 (2001) 1019 1039. [https://doi.org/10.1016/S0048-7333\(00\)00135-9](https://doi.org/10.1016/S0048-7333(00)00135-9)
- O. Gassmann, E. Enkel, Towards a theory of open innovation: three core process archetypes, in: *R&D Management Conference, Lisbon, Portugal, 2004.* <https://doi.org/10.1.1.149.4843>
- R.M. Grant, Prospering in dynamically-competitive environments: organizational capability as knowledge integration, *Organ. Sci.* 7 (1996) 375 387. <https://doi.org/https://doi.org/10.1287/orsc.7.4.375>
- K. Grigoriou, F.T. Rothaermel, Organizing for knowledge generation: internal knowledge networks and the contingent effect of external knowledge sourcing, *Strateg. Manag. J.* 38 (2017) 395 414. <https://doi.org/10.1002/smj.2489>
- M. Gruber, D. Harhoff, K. Hoisl, Knowledge recombination across technological boundaries: scientists vs. engineers, *Manag. Sci.* 59 (2013) 837 851. <https://doi.org/10.1287/mnsc.1120.1572>
- D. Hain, R. Jurowetzki, T. Buchmann, A vector worth a thousand counts: a temporal semantic similarity approach to patent impact prediction, Aalborg, Denmark, Aalborg Universitet 2018. <https://doi.org/2437802164>
- B. Hall, A. Jaffe, M. Trajtenberg, Market Value and Patent Citations: A First Look, 2000. <https://doi.org/10.3386/w7741>
- A. Hargadon, R.I. Sutton, Technology brokering and innovation in a product development firm, *Adm. Sci. Quarterly* 42 (1997) 716 749. <https://doi.org/https://www.jstor.org/stable/2393655>
- D. Harhoff, F. Narin, F. Scherer, K. Vopel, Citation frequency and the value of patented innovation, *Rev. Econ. Stat.* 81 (1999) 511 515. <https://doi.org/https://doi.org/10.1162/003465399558265>
- A. Hatchuel, Y. Reich, P. Le Masson, B. Weil, A.O. Kazakci, Beyond models and decisions: situating design through generative functions, in: *ICED13: 19th International Conference on Engineering Design, Seoul, South Korea, 2013.* <https://doi.org/10.1485144>
- A. Hatchuel, B. Weil, CK Design theory: an advanced formulation, *Res. Eng. Des.* 19 (2009) 181 192. <https://doi.org/10.1007/s00163-008-0043-4>
- A. Hatchuel, B. Weil, A new approach of innovative design: an introduction to C-K theory, in: *International Conference on Engineering Design ICED 03, Madrid, Spain, 2003.* <https://doi.org/citeulike-article-id:4891368>
- C.E. Helfat, Evolutionary trajectories in petroleum firm R&D, *Manag. Sci.* 40 (1994) 1720 1747. <https://doi.org/10.1287/mnsc.40.12.1720>
- R. Henderson, K. Clark, Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms, *Adm. Sci. Q.* 35 (1990) 9 30. <https://doi.org/https://doi.org/10.2307/2393549>
- A. Jaffe, Technological opportunity and spillovers of R&D: evidence from firm's patents, profits, and market value, *Am. Econ. Rev.* 76 (1986) 984 1001. <https://doi.org/https://doi.org/10.3386/w1815>
- A. Jaffe, G. de Rassenfosse, Patent Citation Data in social science research: overview and best practices, *J. Am. Soc. Inf. Sci.* 1 (2010) 2581 2583. <https://doi.org/10.1002/asi>
- S. Kaplan, M. Tripsas, Thinking about technology: applying a cognitive lens to technical change, *Res. Policy* 37 (2008) 790 805. <https://doi.org/10.1016/j.respol.2008.02.002>
- R. Katila, New product search over time: past ideas in their prime? *Acad. Manag. J.* 45 (2002) 995 1010. <https://doi.org/https://doi.org/10.5465/3069326>
- R. Katila, G. Ahuja, Something old, something new: a longitudinal



- study of search behavior and new product introduction, *Acad. Manag. J.* 45 (2002) 1183 1194. <https://doi.org/https://doi.org/10.2307/3069433>
- S.J. Kline, N. Rosenberg, An overview of innovation, *Eur. J. Innov. Manag.* 38 (1986) 275–305. <https://doi.org/10.1108/14601069810368485>
- B. Kogut, U. Zander, Knowledge of the firm, combinative capabilities, and the replication of technology, *Organ. Sci.* 3 (1992) 383 397. <https://doi.org/10.1287/orsc.3.3.383>
- E.D. Kolaczyk, *Statistical analysis of network data : methods and models*, Springer, Berlin, Germany, 2009. <https://doi.org/10.1007/978-0-387-98135-2>
- A. Korotayev, S. Bilyuga, I. Belalov, J. Goldstone, Oil prices, socio-political destabilization risks, and future energy technologies, *Technol. Forecast. Soc. Change* 128 (2018) 304 310. <https://doi.org/10.1016/j.techfore.2017.06.004>
- P. Lane, B.R. Koka, S. Pathak, P.I. Lane, The reification of absorptive capacity: a critical review and rejuvenation of the construct, *Acad. Manag. Rev.* 31 (2006) 833 863. <https://doi.org/https://doi.org/10.5465/amr.2006.22527456>
- K. Laursen, A. Salter, Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms, *Strateg. Manag. J.* 27 (2006) 131 150. <https://doi.org/10.1002/smj.507>
- B. Lawson, D. Samson, Developing innovation capability in organisations: a dynamic capabilities approach, *Int. J. Innov. Manag.* 5 (2001) 377 400. <https://doi.org/10.1142/s1363919601000427>
- P. Le Masson, B. Weil, A. Hatchuel, *Strategic management of innovation and design*, Cambridge University Press, Cambridge, MA, 2010a. <https://doi.org/10.1017/CBO9780511779916>
- P. Le Masson, B. Weil, A. Hatchuel, The design activity and innovation capability, in: *Strategic Management of Innovation and Design*, Cambridge University Press, Cambridge, MA, 2010b. <https://doi.org/10.1017/CBO9780511779916.006>
- D.A. Levinthal, J.G. March, The myopia of learning, *Strateg. Manag. J.* 14 (1993) 95 112. <https://doi.org/0143-2095/93/100095-18>
- L. Leydesdorff, D. Kushnir, I. Rafols, Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC), *Scientometrics* 98 (2014) 1583 1599. <https://doi.org/10.1007/s11192-012-0923-2>
- S. Li, E. Garces, T. Daim, *Technological Forecasting & Social Change Technology forecasting by analogy-based on social network analysis: the case of autonomous vehicles*, *Technol. Forecast. Soc. Chang.* 148 (2010) 119731. <https://doi.org/10.1016/j.techfore.2019.119731>
- J. Lobo, D. Strumsky, Sources of inventive novelty: two patent classification schemas, same story, *Scientometrics* 120 (2019) 19 37. <https://doi.org/10.1007/s11192-019-03102-2>
- J.G. March, Exploration and exploitation in organizational learning, *Organ. Sci.* 2 (1991) 71 87. <https://doi.org/10.1287/orsc.2.1.71>
- C. Martínez, Patent families: When do different definitions really matter? *Scientometrics* 86(1) (2001) 39 63. <https://doi.org/10.1007/s11192-010-0251-3>
- P. Moncada-Paternò-Castello, C. Ciupagea, K. Smith, A. Tübke, M. Tubbs, Does Europe perform too little corporate R&D? A comparison of EU and non-EU corporate R&D performance, *Res. Policy* 39 (2010) 523 536. <https://doi.org/10.1016/j.respol.2010.02.012>
- H. Nakamura, S. Suzuki, I. Sakata, Y. Kajikawa, Knowledge combination modeling: the measurement of knowledge similarity between different technological domains, *Technol. Forecast. Soc. Change* 94 (2015) 187 201. <https://doi.org/10.1016/j.techfore.2014.09.009>
- R. Nelson, G.S. Winter, *An evolutionary theory of economic change*, Harvard University Press, Cambridge, MA, 1982.
- A. Nerkar, Old is gold? The value of temporal exploration in the creation of new knowledge, *Manag. Sci.* 49 (2003) 211 229. <https://doi.org/10.1287/mnsc.49.2.211.12747>
- I. Nonaka, A dynamic theory of organizational knowledge creation, *Organ. Sci.* 5 (1994) 14 37. <https://doi.org/https://doi.org/10.1287/orsc.5.1.14>
- B. Nooteboom, W. Van Haverbeke, G. Duysters, V. Gilsing, A. van den Oord, Optimal cognitive distance and absorptive capacity, *Res. Policy* 36 (2007) 1016 1034. <https://doi.org/10.1016/j.respol.2007.04.003>
- G.C. O'Connor, Major innovation as a dynamic capability: a systems approach, *J. Prod. Innov. Manag.* 25 (2008) 313 330. <https://doi.org/10.1111/j.1540-5885.2008.00304.x>
- C.A. O'Reilly, M.L. Tushman, Ambidexterity as a dynamic capability: resolving the innovator's dilemma, *Res. Organ. Behav.* 28 (2008) 185 206. <https://doi.org/10.1016/j.riob.2008.06.002>
- G. Park, J. Shin, Y. Park, Measurement of depreciation rate of technological knowledge: technology cycle time approach, *J. Sci. Ind. Res.* 65 (2006) 121 127. <https://doi.org/http://hdl.handle.net/123456789/4806>
- R.K. Perrons, How innovation and R&D happen in the upstream oil & gas industry: insights from a global survey, *J. Pet. Sci. Eng.* 124 (2014) 301 312. <https://doi.org/10.1016/j.petrol.2014.09.027>
- C. Phelps, R. Heidl, A. Wadhwa, Knowledge, networks, and knowledge networks: a review and research agenda, *J. Manag.* 38(4) (2012) 1115 1166. <https://doi.org/10.1177/0149206311432640>
- S. Raisch, J. Birkinshaw, G. Probst, M.L. Tushman, Organizational ambidexterity: balancing exploitation and exploration for sustained performance, *Organ. Sci.* 20 (2009) 685 695. <https://doi.org/10.1287/orsc.1090.0428>
- L. Rosenkopf, A. Nerkar, Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry, *Strateg. Manag. J.* 22 (2001) 287 306. <https://doi.org/10.1002/smj.160>
- S Sarica, B. Yan, G. Bulato, P. Jaipurkar, J. Luo, Data-driven network visualization for innovation and competitive intelligence, in: *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Grand Wailea, HI, 2019. <https://doi.org/10.24251/hicss.2019.017>
- A. Schoen, L. Villard, P. Laurens, J.-P. Cointet, G. Heimeriks, F. Alkemade, The network structure of technological developments: technological distance as a walk on the technology map, in: *STI Conference*, Montreal, QC, 2012. 734 742.
- J.A. Schumpeter, *Theory of economic development*, new edition, Routledge, London, 1934.
- M. Squicciarini, H. Denis, C. Criscuolo, Supporting investment in knowledge capital, growth and innovation, OECD, Paris, 2013. <https://doi.org/10.1787/9789264193307-en>
- D. Strumsky, J. Lobo, Identifying the sources of technological novelty in the process of invention, *Res. Policy* 44 (2015) 1445 1461. <https://doi.org/10.1016/j.respol.2015.05.008>
- T.E. Stuart, J.M. Podolny, Local search and the evolution of technological capabilities, *Strateg. Manag. J.* 17 (1996) 21 38. <https://doi.org/10.1002/smj.4250171004>
- J. Sydow, G. Schreyögg, J. Koch, Organizational Path Dependence: opening the black box, *Acad. Manag. Rev.* 34 (2009) 689 709. <https://doi.org/https://doi.org/10.5465/amr.34.4.zok689>
- D.J. Teece, Firm organization, industrial structure, and technological innovation, *J. Econ. Behav. Organ.* 31 (1996) 193 225. [https://doi.org/https://doi.org/10.1016/s0167-2681\(96\)00895-5](https://doi.org/https://doi.org/10.1016/s0167-2681(96)00895-5)
- Thomson-Reuters. Top 100 global energy leaders, 2017. <https://www.thomsonreuters.com/en/products-services/energy/top-100.html>
- M. Trajtenberg, A penny for your quotes: patent citations and the value of innovations, *RAND J. Econ.* 21(1) (1990) 172 187. <https://doi.org/10.2307/2555502>
- D. Verhoeven, J. Bakker, R. Veugelers, Measuring technological novelty with patent-based indicators, *Res. Policy* 45 (2016) 707 723. <https://doi.org/10.1016/j.respol.2015.11.010>
- R. Veugelers, J. Wang, Scientific novelty and technological impact, *Res. Policy* 48 (2019) 1362 1372. <https://doi.org/10.1016/j.respol.2019.01.019>
- J. West, M. Bogers, Leveraging external sources of innovation: A review of research on open innovation, *J. Prod. Innov. Manag.* 31 (2014) 814 831. <https://doi.org/10.1111/jpim.12125>
- J. West, A. Salter, W. Vanhaverbeke, H. Chesbrough, Open innovation: The next decade, *Res. Policy* 43 (2014) 805 811. <https://doi.org/10.1016/j.respol.2014.03.001>
- WIPO, *Guide to the International Patent Classification*, World Intellectual Property Organization, Geneva, Switzerland, 2019.
- S. Yayavaram, G. Ahuja, Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability, *Adm. Sci. Q.* 53 (2008) 333 362.
- S. Yayavaram, W.–R. Chen, Changes in firm knowledge couplings and firm innovation performance: the moderating role of technological complexity, *Strateg. Manag. J.* 36 (2015) 377 396. <https://doi.org/10.1002/smj>

**Quentin Plantec** is a Ph.D. candidate at the Centre for Management Science (CGS) — i3 (UMR CNRS) at MINES ParisTech — PSL University, Chair of Design Theory and Methods for Innovation (DTMI). He holds an MS from Ecole Polytechnique and from Ecole Normale Supérieure de Cachan in Economics & Business. He has been awarded a 3-year grant by the Institut National de la Propriété Industrielle (French Patent Office – CIFRE scheme) to support research works. His research focuses on innovation, science–industry links and patent analytics. He has been published in conference proceedings of R&D Management, European Academy of Management (EURAM) and International Conference on Engineering Design (ICED).

**Pascal Le Masson** is a Professor at MINES ParisTech — PSL University, Chair of Design Theory and Methods for Innovation (DTMI). He is a part of the management team of the Center for Management Science — i3 (UMR CNRS) and a co-head, with Benoit Weil, of the Engineering Design curricula. He conducts research on Design Theory and Methods for Innovation, with five primary topics: design theory; governance and methods for the innovative firm; neuroscience of creative design; innovative ecosystems organization and regulation; object identity. He has published “Strategic Management of Innovation and Design” (Cambridge University Press) and “design theory” (Springer) (both co-authored by Armand Hatchuel and Benoit Weil), with several papers. He conducts collaborative research with several companies (start-ups, SMEs and corporate firms), in particular with the partners of the DTMI Chair.

**Benoit Weil** is a Professor at the Center for Management Science–i3 (UMR CNRS 9217), MINES ParisTech, France. He is also the Chair of Design Theory and Methods for Innovation (DTMI) and the Head of the Engineering Design curriculum of MINES ParisTech. His research focuses on the rationalization of collective actions. He has created the Research Program on Design Activities with Professor Armand Hatchuel. Together they proposed a new theory of design reasoning (C–K theory), which accounts for the dual expansion of knowledge and concepts characteristic of innovative design. He has published several papers and books.

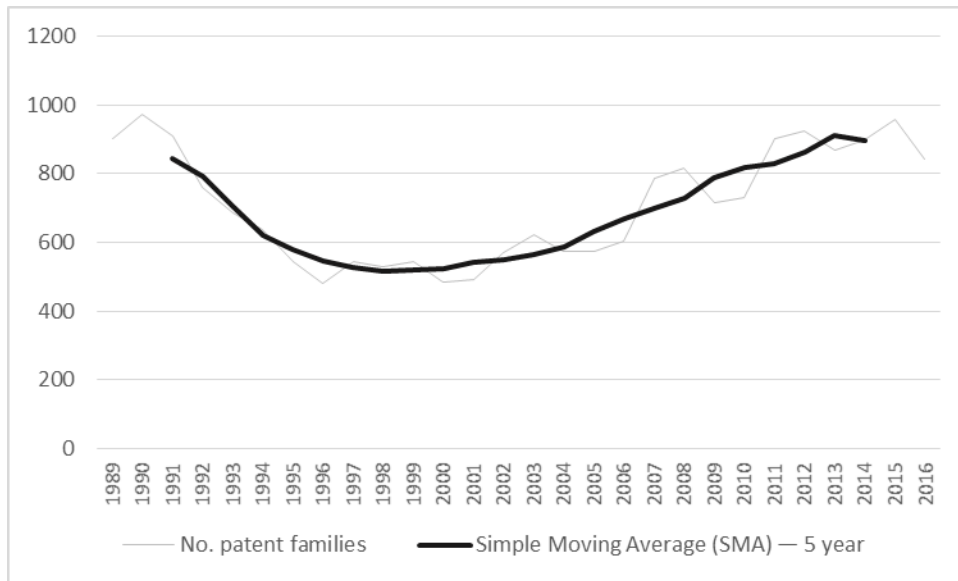
## 9. APPENDIXES

Company	Nationality	No. patent families	Breakdown no. patent families per decade				No. utility patents
			1980 — 1989	1990 — 1999	2000 — 2009	2010 — 2019	
BP	United Kingdom	5,406	2,509	1,592	816	489	12,437
Chevron	United States of America	5,598	2,374	1,073	1,060	1,091	13,281
ConocoPhillips	United States of America	4,060	1,991	1,107	505	457	8,492
Eni	Italy	1,907	387	679	447	394	5,217
Equinor	Norway	823	46	170	273	334	1,876
Exxon Mobil	United States of America	14,113	5,389	3,681	2,693	2,350	40,122
Indian Oil Corp.	India	173	-	15	37	121	350
INPEX Corp.	Japan	93	1	2	33	57	363
PetroChina	China	260	-	5	58	197	438
Reliance Industries	India	360	4	10	52	294	841
Repsol	Spain	185	-	51	44	90	436
Saudi Basic Industries Corp.	Saudi Arabia	2,995	14	91	727	2,163	8,029
Royal Dutch Shell	Netherlands	8,793	2,505	2,218	2,125	1,945	22,874
SK Innovation Corp.	Korea	680	-	-	167	513	1,787
Suncor Energy	Canada	123	37	18	28	40	286
Total Group	France	3,093	388	640	871	1,194	9,359
<b>Total</b>	-	<b>48,662</b>	<b>15,645</b>	<b>11,352</b>	<b>9,936</b>	<b>11,729</b>	<b>126,188</b>

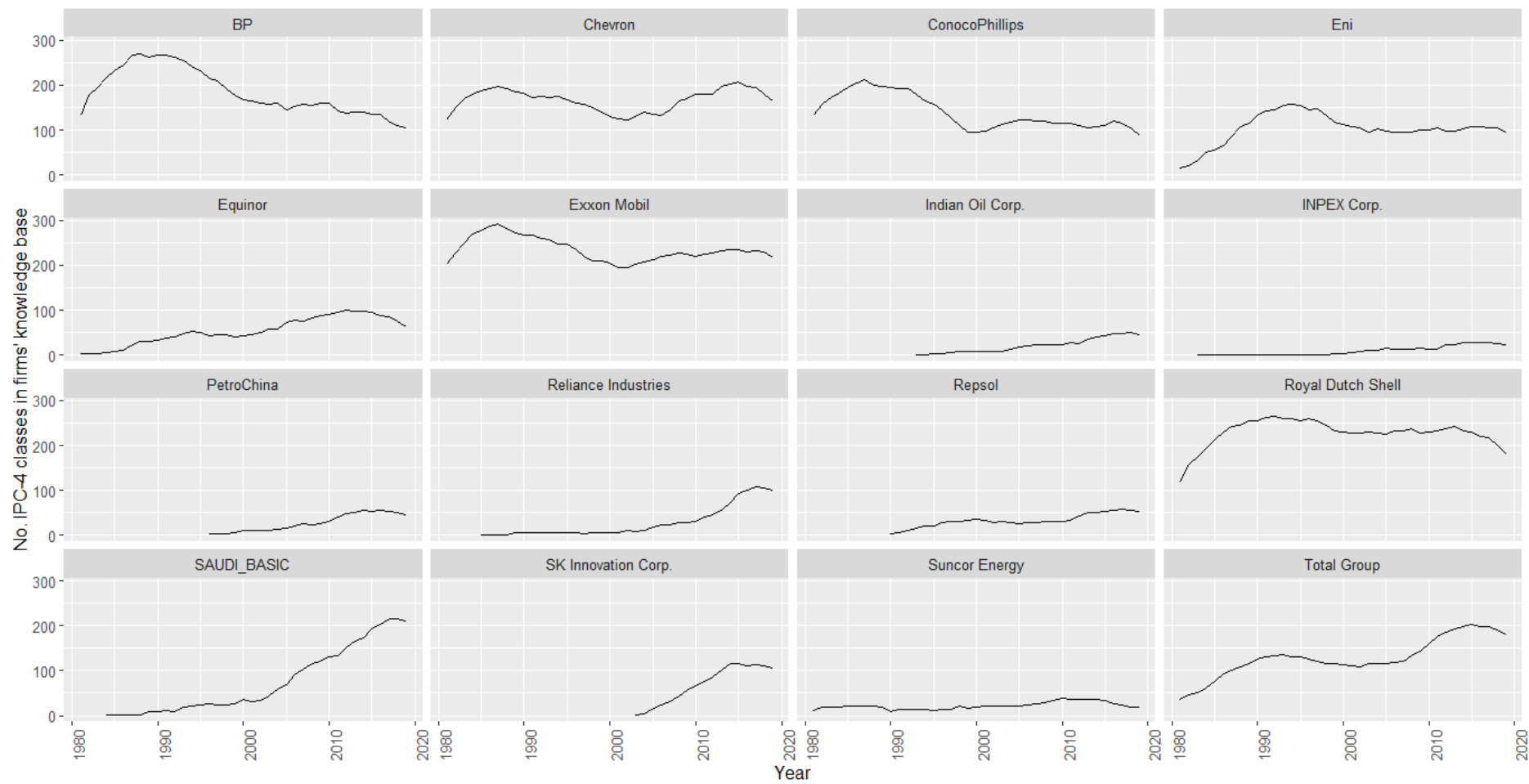
Appendix 1 - Original data set breakdown by firm

IPC-4 Class ID	Description	Occurrence
B01J	CHEMICAL OR PHYSICAL PROCESSES (CATALYSIS, COLLOID, etc.)	5,885
C07C	ACYCLIC OR CARBOCYCLIC COMPOUNDS	4,607
C10G	CRACKING HYDROCARBON OILS; PRODUCTION OF LIQUID HYDROCARBON MIXTURES	3,996
C08L	ORGANIC MACROMOLECULAR COMPOUNDS	2,664
C08F	MACROMOLECULAR COMPOUNDS OBTAINED BY REACTIONS ONLY INVOLVING CARBON-TO-CARBON UNSATURATED BONDS	2,400
B01D	CHEMICALS OF PHYSICAL PROCESSES OF SEPARATION	1,916
E21B	EARTH OR ROCK DRILLING	1,883
C10M	LUBRICATING COMPOSITIONS	1,646
C01B	NON-METALLIC ELEMENTS (FERMENTATION, ELECTROLYSIS, ELECTROPHORESIS)	1,615
C08G	MACROMOLECULAR COMPOUNDS OBTAINED OTHERWISE THAN BY REACTIONS ONLY INVOLVING CARBON-TO-CARBON UNSATURATED BONDS	1,383
C10L	FUELS NOT OTHERWISE PROVIDED FOR; NATURAL GAS; SYNTHETIC NATURAL GAS OBTAINED BY PROCESSES	1,379
C08K	USE OF INORGANIC OR NON-MACROMOLECULAR ORGANIC SUBSTANCES AS COMPOUNDING INGREDIENTS	1,359
C08J	WORKING-UP; GENERAL PROCESSES OF COMPOUNDING	1,181
C07B	GENERAL METHODS OF ORGANIC CHEMISTRY	1,067
C09K	CHEMISTRY MATERIALS FOR APPLICATIONS NOT OTHERWISE PROVIDED FOR	1,049
C10N	PETROLEUM, GAS, OR COKE INDUSTRIES; TECHNICAL GASES CONTAINING CARBON MONOXIDE; FUELS; LUBRICANTS; PEAT	984
G01V	GEOPHYSICS; GRAVITATIONAL MEASUREMENTS; DETECTING MASSES OR OBJECTS	786
B32B	LAYERED PRODUCTS, i.e., PRODUCTS BUILT-UP OF STRATA (FLAT OR NON-FLAT)	709
G01N	ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES	686
C07D	HETEROCYCLIC MACROMOLECULAR COMPOUNDS	682

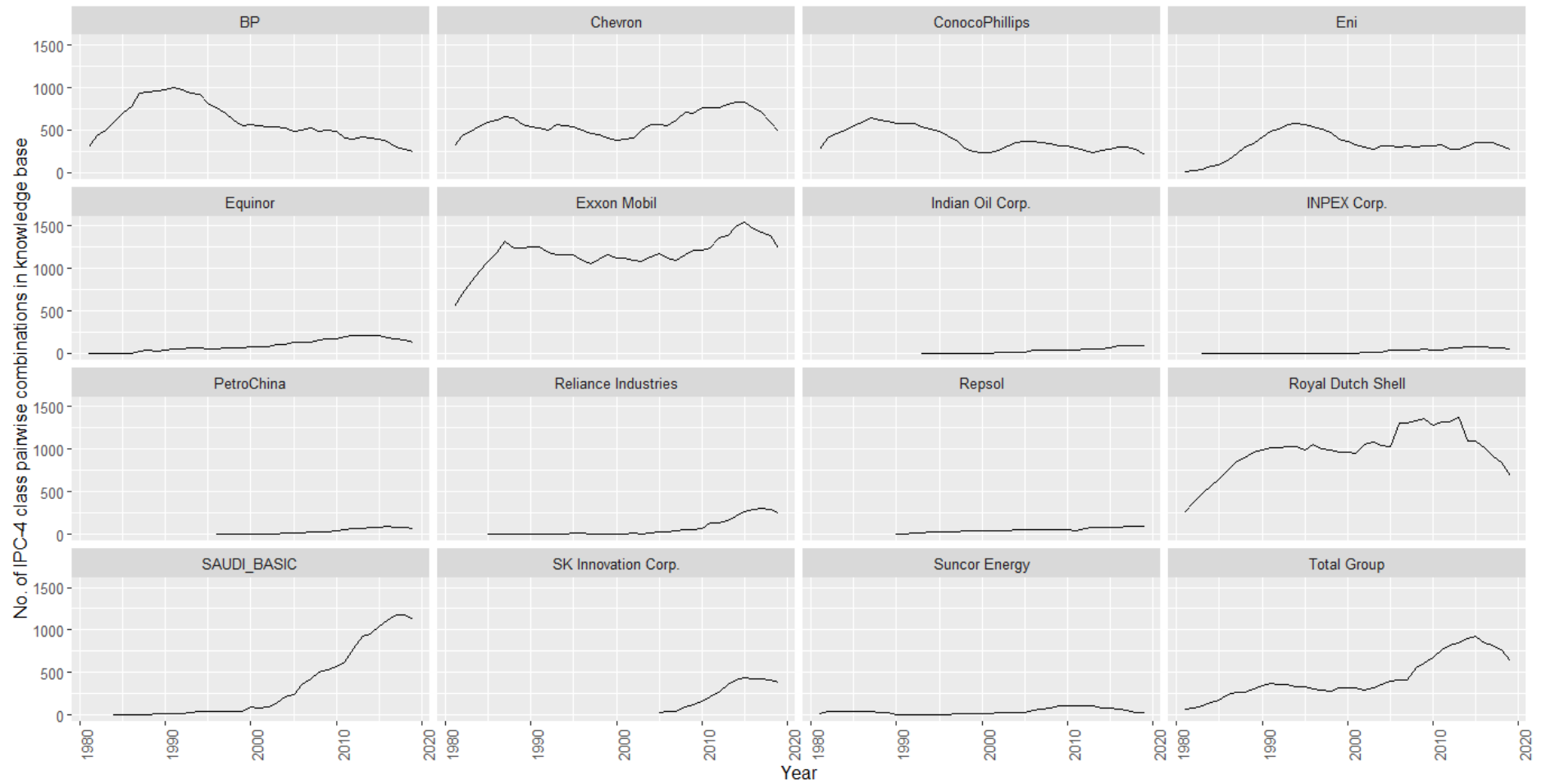
Appendix 2 - Top 20 IPC-4 classes and description in the sample



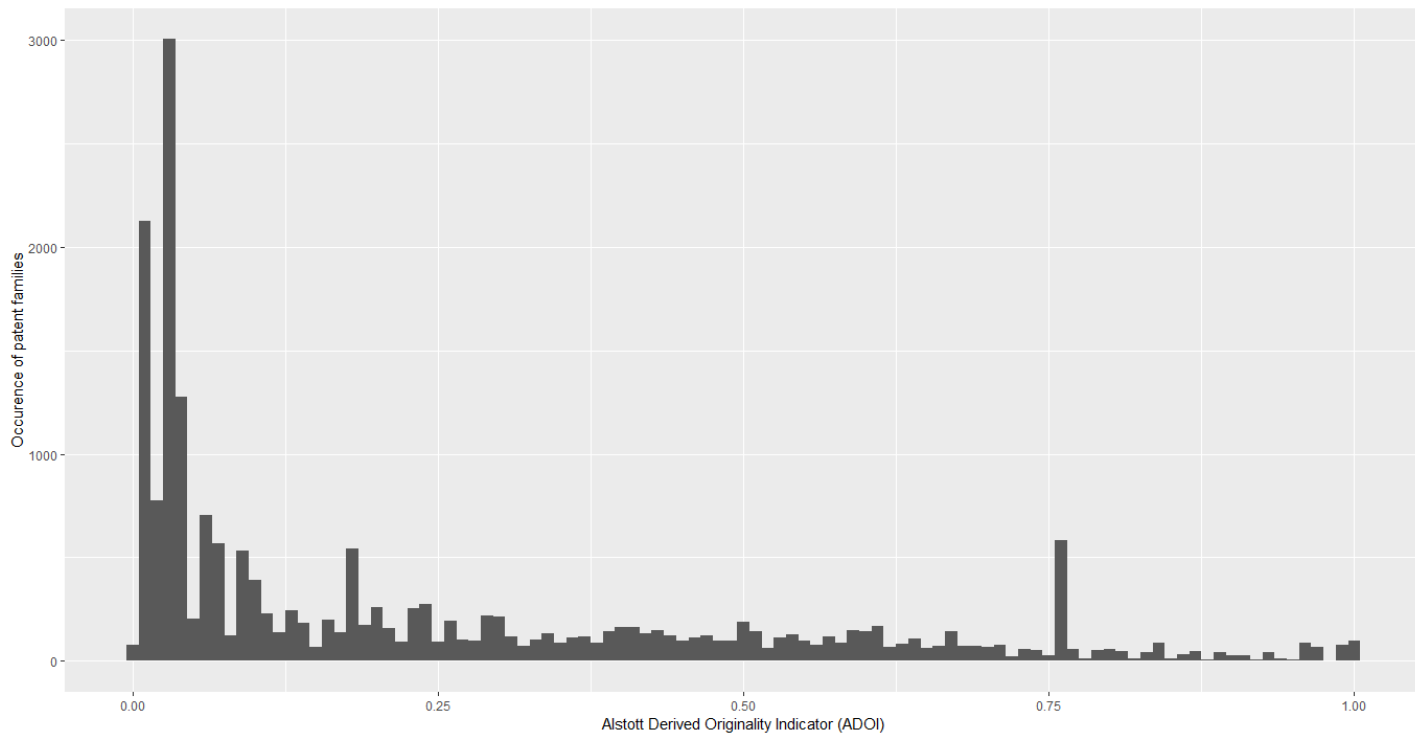
Appendix 3 - Number of INPADOC patent family applications in the sample per year



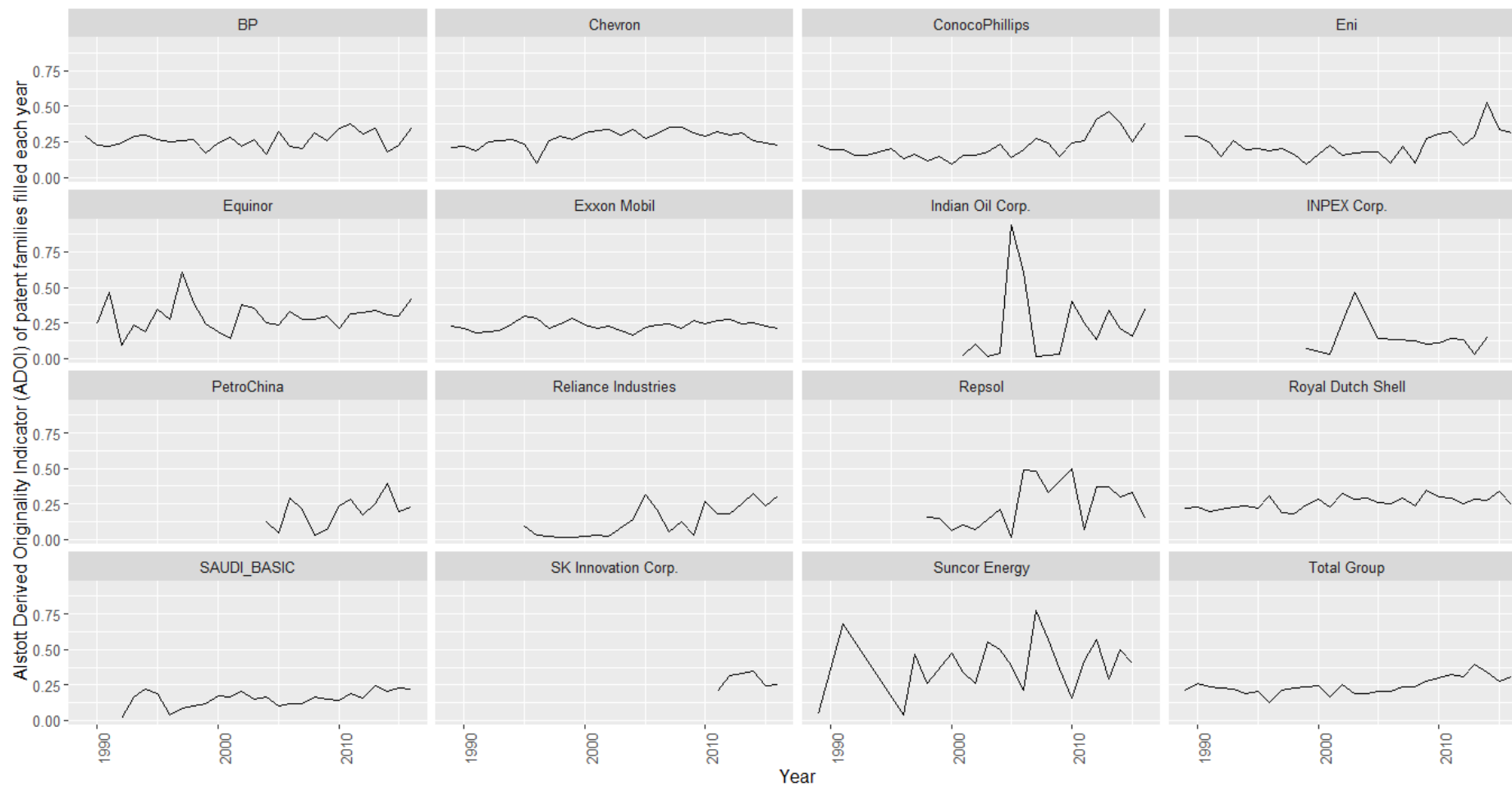
Appendix 4 - Number of IPCs per year in firm's knowledge base



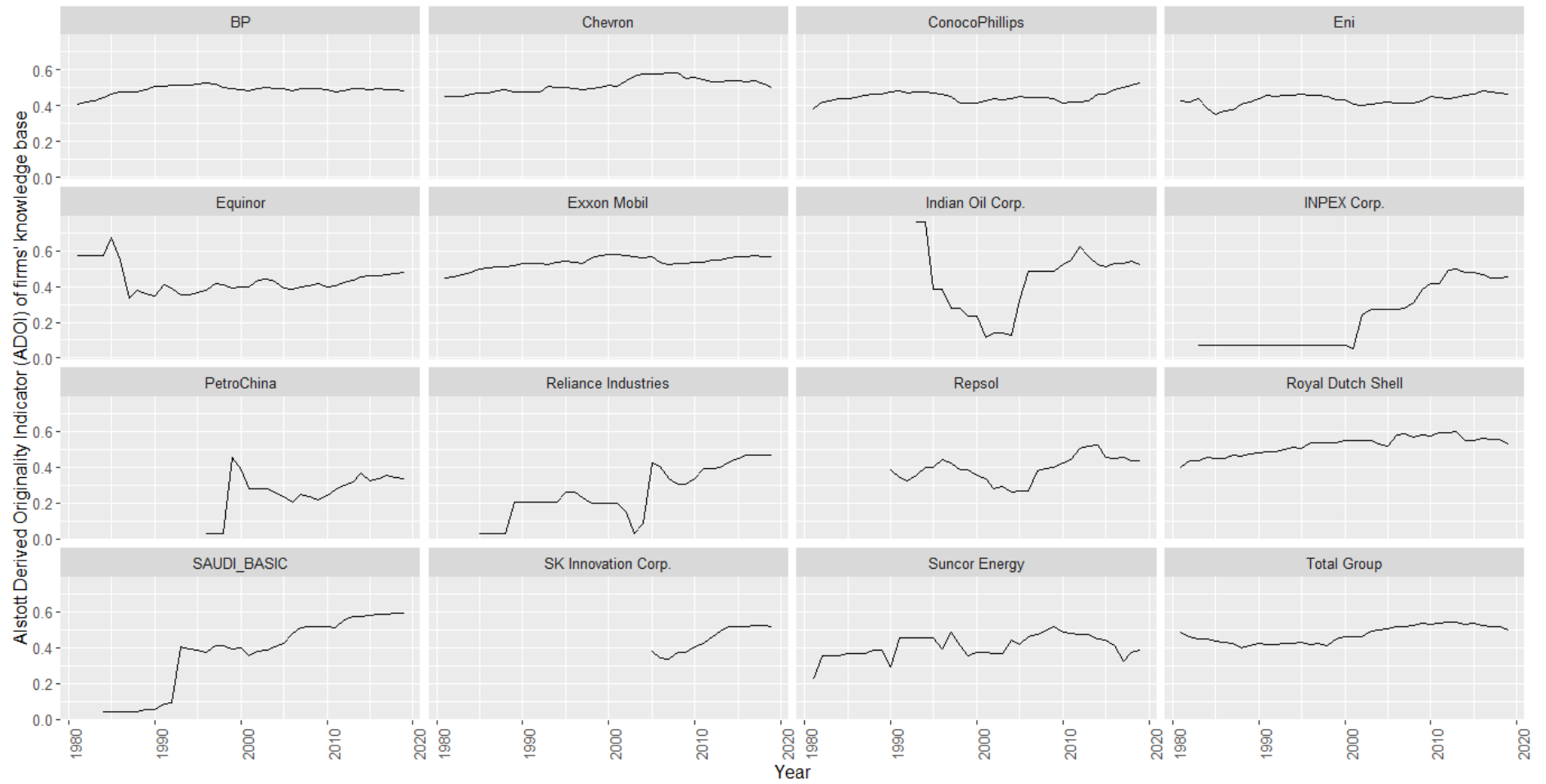
Appendix 5 - Number of IPC pairwise combinations in firm's knowledge base



Appendix 6 - ADOI histogram for the sample



Appendix 7 - ADOI of yearly patent family application average



Appendix 8 - ADOI applied to firm's knowledge base



	ADOI of recombination mode inventions	95% confidence interval		Occurrence	Share of recombination mode	p-value
		min.	max.			
Recomposition mode - without new knowledge components	0.448	0.427	0.469	528	76.6%	p < 0.001
Recomposition mode - with new knowledge components	0.525	0.494	0.555	161	23.4%	p < 0.001
<b>Recomposition Mode - average</b>	<b>0.466</b>	<b>0.448</b>	<b>0.484</b>	<b>689</b>	<b>100.0%</b>	<b>p &lt; 0.001</b>

Appendix 9 - Analysis ADOI score and recombination mode—details