



Of leaders and laggards - Towards digitalization of the process industries

Lukas Jan Aaldering^a, Chie Hoon Song^{a,b,*}

^a Institute of Business Administration, Department of Chemistry and Pharmacy, University of Münster, Leonardo Campus 1, 48149, Münster, Germany

^b REACH – EUREGIO Start-up Center, University of Münster, Geiststraße 24, 48151, Münster, Germany

ARTICLE INFO

Keywords:

Digitalization
Information technology
Market scouting
Network
Process industries
Startup

ABSTRACT

The digital wave of change had an unprecedented effect on the competitiveness of the global value chain, whereby the process industries are not an exception. Although they are often regarded as laggards in the general progress towards adopting digital technologies, the pressure to cut costs and to review existing business models have prompted the industry forerunners within the process industries to reassess their business operations. One way to tackle this issue is to source information from the emerging startup landscape. To shed light on the importance of digitalization within the process industries, this paper provides a novel perspective using startup information. Therefore, it examines the interaction dynamics between the process industries and other industries. By centering on emerging interactions with information technology (IT)-related industries, this study further focuses on identifying the different digitalization efforts of the process industries and explores their varying degree of digital affinity. Our results indicate that not all process industries can be regarded as laggards in terms of incorporating digital capabilities. “Biotechnology”, “Pharmaceutical”, “Food and Beverage”, “Energy” and “Oil and Gas” demonstrated a higher IT-affinity, thus presenting themselves as digital leaders within the process industries. Judging from the IT-interaction preference, each segment of the process industries has adopted a unique pathway towards unlocking digital transformation opportunities.

1. Introduction

An important aspect in today's business environment is the diffusion of new industrial paradigm characterized through merging of automation and data exchange in manufacturing, commonly referring to Industry 4.0 (Castelo-Branco et al., 2019; Lasi et al., 2014). Industry 4.0 denotes a transformation from machine dominant manufacturing to digital manufacturing and bears a huge potential to achieve a higher level of personalization and resource efficiency (Benitez et al., 2020; Diez-Oliván et al., 2019). It is safe to say that there is no industry remaining untouched by this digital revolution (Mariani and Borghi, 2019). The amount of available data has continuously increased over the past years in a non-linear fashion, and new analytical approaches combined with greater computational power are turning this flood of raw data into meaningful insights (Oztemel and Gursev, 2020; Xu and Duan, 2019). In product or service innovation, the application of advanced analytics, such as machine learning or natural language processing, is transforming the way companies base their decisions and interact with the value chain partners (Gosal et al., 2019). Many companies have already harnessed these capabilities to improve their core

operations or to rethink existing business models (Alejandro Germ Frank et al., 2019). For example, leading e-commerce companies or media-services providers are personalizing their customer experience with artificial intelligence (AI)-powered product recommendations (Zhang et al., 2019). In terms of manufacturing performance, a greater openness to enabling technologies of Industry 4.0 helps companies achieve higher speed in prototyping and increased production capacity due to the possibility of demand estimations (Büchi et al., 2020). In view of sustainability, new opportunities arise to lower waste from production due to the possibility of receiving information in real-time and providing virtual training (Brozzi et al., 2020).

Against this background, the pressure to find strategies for cost reduction and to review business models have prompted the industry forerunners within the process industries to reevaluate their business operations (Snihur and Wiklund, 2019). The cluster of process industries is a part of all manufacturing industries and spans multiple industrial segments, including petrochemicals, chemicals, food and beverages, mining and metals, mineral and materials, pharmaceuticals, pulp and paper, steel and utilities (Lager, 2017). As more information will be available to manufacturing plants and extended supply chains, it is

* Corresponding author. REACH – EUREGIO Start-up Center, University of Münster, Geiststraße 24, 48151, Münster, Germany.

E-mail address: chiehoon.song@uni-muenster.de (C.H. Song).

expected that Industry 4.0 will lead to major changes in the configuration of manufacturing companies in terms of their value contribution and production network (Culot et al., 2020). The future smart manufacturing systems will have a different vision of interconnected product, supplier intelligence, communication and information network to better respond to customer needs and disruptions in the market place (Oztemel and Gursev, 2020). The smartness will go beyond total automation and control of the systems (Kusiak, 2018).

However, analyzing how the process industries are adopting this digital transformation is challenging, given that their innovations are somewhat dependent on the external provision of machines, equipment and software. Typically, the manufacturing industries are customers of high-tech innovators, whereby the improvements in productivity result from the utilization of embodied technology purchased from other sectors (Trott and Simms, 2017). Unfortunately, there is no objective measure to determine such activities in the formal statistical measurements used for research and development (R&D). Moreover, there was an image that ICT-induced productivity effects are less pronounced in manufacturing industries (International Business Publications, 2009). This may be also linked to the fact that the related innovative activities take place within a production plant and can only be captured through learning-by-doing and extensive tacit knowledge (Trott and Simms, 2017). This makes more difficult to determine the influence of digital transformation using conventional data sources.

On the other hand, emerging startups are capitalizing on the disruptive trends of Industry 4.0 by transcending sectoral demarcation, often having analytics- or digital-related backgrounds (Lim et al., 2018). Startups refer to newly founded companies, which are perceived as a potentially rich source of novel ideas for bringing innovations to the market and are flexible in exploring untapped business opportunities (Mercandetti et al., 2017). Compared to established companies, they are in a better position to make use of their low level of bureaucracy, high agility and openness to come up with targeted and innovative digital solutions for traditional businesses that rely on physical assets (Nambisan, 2017). Due to such characteristics, they can act as a useful data source for technology or market scouting in the era of digital transformation (Rindfleisch et al., 2017). Nevertheless, studies dealing with the influence of digital capabilities on the process industries from a startup-based perspective are scant. Thus, the aim of this study is to analyze whether individual segments of the process industries have unique requirements in terms of adopting digital technologies and whether a different approach is required when implementing Industry 4.0 through the lens of startup information. This study is conducted based on exploratory research, aiming to identify patterns and relationships between the segments of the process industries and segments of digital transformation. Thus, it attempts to provide sophisticated perspectives on the opportunities arising from the digital transformation for the process industries and to make contributions towards emergence of new knowledge and theories in the process industries discipline. Particular attention is given to analyzing and visualizing the different degree of interaction between the segments pertaining to the process industries and the segments representative for digital transformation. Consequently, three research questions are formulated as follows:

RQ1: Can startup information be used as a means for investigating the influence of digital technologies on the process industries?

RQ2: Which industries act as drivers for digital transformation?

RQ3: Do the individual segments of the process industries behave differently when adopting digital technologies and if so, what opportunities arise for these sectors?

To answer these questions, we performed data-driven analyses focusing on the investigation of interaction dynamics between industries using data drawn from the startup landscape. After providing general statistics about the development trend of startups within the process industries, we quantified and visualized the interacting pairs of industries by means of network representation. Additionally, the degree of IT-affinity and unique IT-interaction preference were examined by

means of heat map representation. Compared to previous studies, which predominantly rely on the qualitative research interview, this study can help reveal emerging patterns that may occur in the future or may not be known to industry representatives or R&D managers. Hence, in terms of theoretical contribution, this study complements the survey- or interview-based studies by providing a novel perspective to investigate the digital transformation. In terms of managerial contribution, the findings help the decision makers to better understand the chances arising from the digitalization and to reveal vacant digitalization opportunities that could be exploited to strengthen the integration process of digital solutions.

The remainder of this study is organized as follows: Following this introduction, Section 2 briefly reviews the theoretical background of process industries, startups and their use as a data source for technology and market scouting. Section 3 outlines the data collection procedure and the applied methodological approaches. In Section 4, the major research findings are presented. Section 5 discusses the findings with reference to the research questions of this study. In Section 6, the managerial and theoretical implications are presented. The final section summarizes the major findings and concludes with an outlook for future research.

2. Background

2.1. Process industries in upheaval

The process industries were originally defined as a cluster of various industries spanning chemicals, petrochemicals, pharmaceuticals, mineral, mining, metal, food processing, pulp and steel (Lager, 2010). Recently, this definition was extended to represent “a part of all manufacturing industries, using raw-materials (ingredients) to manufacture non-assembled products in an indirect transformational production process often dependent on time” (Lager, 2017). Hence, the cluster of process industries also includes biotechnology, energy (e.g. electricity), and food and beverage industries. They share common characteristics concerning the production process, thus making up a substantial part of the manufacturing industry, and are characterized by a complex supply chain, high asset intensity and high degree of attachment to production sites (Lager et al., 2013). A major difference between the process industries and other manufacturing industries lies in the type of incoming and outgoing material (Lager, 2010). Primarily, the process industries are responsible for processing raw materials and delivering intermediate products for the downstream market rather than manufacturing of prefabricated components or assembled products. Moreover, the different unit processes along the long supply chain are connected in a continuous flow. Thus, the successful development of a new process technology or product innovation requires a profound understanding of the total chain operation. In this context, it has been argued that there is little room for responding to dynamically changing market conditions (despite the presence of several supplier-customer interfaces), as the ability of process firms to reach the end users as well as to alter the marketing perspectives is limited (Hill, 2000; Zaborek and Mazur, 2019).

Traditionally, an improved innovation performance can be achieved through intense collaboration with providers of equipment suppliers. New product concepts within the process industries commonly require changes to the production process and subsequently the process development also evokes changes in features of the final product (Frishammar et al., 2012). Such collaboration allows process firms operating in more mature industries to acquire competences or skills that would otherwise not have been accessible, whereby equipment suppliers can use this relationship to gain feedback on new prototype technology. Hence, innovation opportunities are more easily missed, if there are no mutual commitment to find competitive production solutions from both sides. In order to effectively respond to competitive pressure, firms in the process industries are suggested to deal with dynamics of the total cost

structure in the production process, thereby developing innovation strategies that embrace both the product-market and the process-manufacturing perspective (Onufrey and Bergek, 2020; Storm et al., 2013). Especially, for firms producing more functional products rather than commodities, the pursuit of application development, which focuses on optimizing cost saving potential through involvement in the customer's process and product development, can add value to the customer's production economy and secure an opportunity to build long-term sustainable customer relationships (Lager and Storm, 2013). In a recent study, a framework for platform-based production and design of non-assembled products within the process industries was developed to improve management of product variety under operational constraints (Samuelson and Lager, 2019). Using a such framework, both economies of scale and scope can be maintained whilst serving as a tool for assessing corporate strategic production capabilities.

Today, the process industries have been hit by advancements in the digital transformation, which have resulted in a substantial need for cross-disciplinary competences and research collaborations to stimulate innovation (Isaksson et al., 2018). Machine learning, big data and connected devices are shaking up multiple industries, and the impact will only become more pronounced, as their acceptance reaches the critical mass (e.g. allowing machines to mimic and surpass human intelligence) (Yao and Ge, 2018) (Note: A glossary is attached in Appendix for clarifying the definition of technical terms used, such as "machine learning" and "big data"). New opportunities for value creation are expected to arise from incorporating networked machines and data-based operating models into a firm's strategic agenda. In fact, the process industries can use insights derived from huge volumes of data generated via devices and sensors to make faster evidence-based decisions (Stock and Seliger, 2016). For example, the chemical industry can use "data" as another type of raw material and establish new digital business models that can act as aggregation platforms between end users and business ecosystem participants. Moreover, digitalization can automate the production process with regard to optimizing the resource input and finding bottlenecks. It can also help correct errors that interrupt the workflow. However, progress in capturing value from data and predictive analytics has been not even over the industrial landscape. So far, the process industries have captured only a fraction of potential values compared to their peer industries (e.g. retail or automotive industry) that take a leading role in the push for digital sovereignty (Reinartz et al., 2019). Although advances in data science and the convergence of several digital technology trends can allow process firms to tackle business problems with a whole new mindset and outlook, factors, such as lack of senior management involvement and the design of right organizational structure and technological infrastructure (e.g. perception that the complexity outweigh the potential gains), were identified as critical internal barriers (McKinsey Global Institute, 2016).

Consequently, it is generally assumed that organizations capable of harnessing digital capabilities will create unique value and differentiate themselves, while others will find themselves increasingly at a disadvantage. Herein, the demand for business translators who can act as a broker between analytics and practical applications to business questions would increase in the future. Simultaneously, firms need to reinforce its capability to network and tap into the resources of startups, as the diversity of network partners is important for transforming the corporate culture and fully exploiting the opportunities arising from digital-driven economic networks. This implies that firms need to take a broad ecosystem perspective to stimulate innovation and draft their strategic development.

2.2. Startups as a driver behind disruptive forces

Research on entrepreneurship has gained momentum in the last two decades (Cavallo et al., 2019; Saebi et al., 2019). The emergence of novel digital technologies and associated changes in digital infrastructures has transformed the entrepreneurial business environment, stimulating new

forms of innovation and entrepreneurial initiatives that cross-link traditional sectoral boundaries (Nambisan et al., 2019). It has been noted that the entrepreneurial ecosystems are characterized by high degree of digital affordance, emphasis on business model innovation and horizontal knowledge spillovers (Autio et al., 2018). Hence, it might be useful to view modern entrepreneurship as part of digital economy phenomenon that harnesses technological advances to be actively engaged in the redesign of value creation, delivery, and capture processes by novel ventures (Nambisan et al., 2019).

In general, startups refer to newly founded companies with limited operating history but are flexible in exploring untapped business opportunities and providing novel solutions to problems which are rather unknown in a volatile market environment. They do not only act as engines of job creation but also contribute to disruption of existing industry paradigms and identification of services gaps (Decker et al., 2014). The transformative trends of digitalization have lowered the entry barriers for new ventures, whose business model builds upon an agile, asset-light, and technology-driven infrastructure (Deloitte, 2017). This enabled them to have an exponential growth even with a very lean setup. However, the lack of resources or initial capital forces startups to collaborate with larger firms, whereby their opportunity seeking behavior allows them to transfer their expertise in order to successfully launch new offers or to better connect the demand with supply side. Moreover, startups are often created to develop technology that is seen as cutting edge, and strive for building a scalable business that grows in a very fast and efficient way. Due to this exploratory and interactive nature of startups, it is of great interest to investigate the startup landscape, as they tend to be more open and adaptable towards shifting customer expectations (Antolín-López et al., 2015; Massa and Testa, 2008). Consequently, startups play an important role in introducing industrial values geared towards sustainable development of market economy and further drive traditional manufacturing companies to digitize their functions and processes (Stayton and Mangematin, 2016).

2.3. Startups as a new data source for technology and market scouting

In terms of analyzing the technological landscape and associated business environment, scholars have predominantly relied on the use of patent statistics (Huang and Su, 2019; Wu and Mathews, 2012). Patent data serve as a unique source of information on technological innovation, as they are issued by a government authority and represent a direct outcome of inventive activity having a commercial outlook. They contain standardized information relevant for making strategic decisions and permit monitoring of technological change as well as evaluating the impact of patent-based indicators on firm performance (Verhoeven et al., 2016). Moreover, studying the technological inter-disciplinarity or interdependence can help disclose the emerging research clusters and the characteristics of strongly interacting industries (Breitzman and Thomas, 2015). Herein, the use of patent classification has found wide applications in exploring the changing boundaries of industrial ecology (Curran and Leker, 2011; Lee et al., 2019).

However, patent data cannot entirely cover all inventions or disruptive market innovations, especially those which are technically not patentable or protected in an alternative way. In particular, the analysis of technology-service convergence, which is driven by the digital union of new generation of products and services generated from cross-linking of disparate industries, is not an easy task. Rapidly changing market dynamics cannot be adequately spotted using patent information only (Kodama, 2014). Although data related to mergers and acquisitions (M&A) transactions or strategic alliances can be used to quantify and visualize the competitive market dynamics shaping the business ecosystem (Aaldering et al., 2018; Basole et al., 2015), these data are more suitable for describing inter-firm linkages in established businesses where changes in knowledge couplings are manageable. New industries that have their origin in the digital revolution (such as "Apps"

and “Cloud Computing”) require differentiated data categories that better reflect the state-of-the-art in the market space. For example, in case of SIC (Standard Industrial Classification) classification, mobile apps would be classified in the same category as software application. Digital entrepreneurs often put incumbents at risk by unlocking new business value through data relationships and are characterized by different streams of knowledge. Therefore, the application of existing data (e.g. SIC or patent data) could be difficult to gain insights into boundary-spanning exploration of coupling between new and existing knowledge domains (Basole et al., 2019).

Against this background, startups can act as an alternative data source for technology (or market) scouting, which helps detect and integrate advances in technology (or market) trends at an early stage. Technology scouting encompasses a systematic process of identifying and evaluating emerging technologies as well as facilitating their acquisition (Rohrbeck, 2010). In the age of digital disruption, the spectrum of possible data sources must be extended to manage uncertainties and to exploit service-side innovation. As startups (or new technology ventures) play a significant role in boosting the innovation potential of incumbents (Becker et al., 2016) and enhancing the social wealth of communities (Abu-Saifan, 2012), they can act as a useful data source for identifying new development opportunities within the industries seeking to adapt to the waves of digital transformation or to work with partners with less conventional capabilities.

3. Research design and methodologies

3.1. Data collection

Crunchbase is a crowdsourced database for finding business information about new ventures and related legal developments. It is considered as a leading source of intelligence on technology-based startups and provides competitive insights into funding history and acquisition of startup companies (Ko and McKelvie, 2018). Although Crunchbase relies on a set of external investor networks and user contributions to update and maintain the database, the data integrity and validity are achieved through application of machine learning algorithm capable of reviewing data accuracy and a team of experts, who are responsible for manual data validation and curation (Marra et al., 2017). Due to this unique approach to data collection, data contained in Crunchbase are characterized by an added value compared to other commercial databases and public data sources (Marra et al., 2020). To benchmark the data coverage of Crunchbase, Dalle et al. (2017) compared it with the OECD Entrepreneurship Financing Database, and found similar coverage patterns across two data sources. Furthermore, Crunchbase has been used as a reliable data source in over 90 scientific publications and many scholars have recognized its potential for advancing new lines of economic and managerial research (Dalle et al., 2017). In the present study, we used Crunchbase to find startups that belong to the process industries and were established between 2010 and 2018. We adopted this time frame, as it allows us to incorporate new ventures that are in early stages of their financing. The probability of a startup to be registered at the database is affected by the public awareness level of a startup company. Hence, the number of actual startups for the year 2019 may vary significantly from the actual values. To provide a more reliable picture of the startup landscape, we decided to include only data up to 2018. Moreover, we can avoid the full brunt of financial crisis, which might have negative effects on venture financing, by setting the start year to 2010.

The database provides information on, but is not limited to, location of headquarter, number of employees, industry category codes, total amount of funding raised, number funding rounds, and other metadata related to products or services. To ensure the focus on the process industries, we used a combination of industry category codes to limit the initial search results. In analogy to the Standard Industrial Classification (SIC) codes, industry category codes are a classification scheme

provided by Crunchbase that labels startups according to capabilities and knowledge on which companies build their own specializations. Subsequently, a startup can have several core competencies, thus being characterized by multiple industry category codes. Startups with industry category codes such as “Chemical” and “Machine Learning” would imply that it runs and organizes their business activities using the expertise (or technologies) from both fields. In the following, we will use the term *industry* interchangeably with *industry category code* to denote the segment of economic activity that produces goods and services based on the application of similar knowledge and skills. To filter out relevant startups, we selected entries with industry category codes that contain at least one of the following terms: “Biotechnology”, “Chemical”, “Pharmaceutical”, “Food and Beverage”, “Energy”, “Metal and Mining”, “Oil and Gas”, “Paper Manufacturing” and “Wood Processing”. This is in line with the definition provided by (Lager, 2017). A total of 16,429 startup companies were included in the sample data.

3.2. Exploration of interacting industries

To demonstrate the suitability of using startup information as a proxy for examining the impact and reach of digitalization trends in the process industries, we employed a data-driven exploratory approach. This exploratory approach aims to lay groundwork for further research on innovation management in the process industries by examining the emerging startup landscape. For this purpose, the concept of knowledge flow was adopted to quantify and visualize the interaction dynamics among different industry category codes (Song et al., 2017). In general, knowledge flows can be regarded as the exchange of technological or market knowledge within and across organizational boundaries, which contribute to enhanced problem solving by enlarging the organization’s knowledge stock (Argote and Ingram, 2000). They can be either directed or undirected, depending on whether there exist directional dependencies between the knowledge carriers. Based on the phenomena of knowledge flow, questions regarding the origin of knowledge source or identification of interdisciplinary technology areas can be answered (Cavaggioli, 2016; Singh, 2005). As innovation through the diffusion of knowledge has become a major driver behind economic growth, understanding the patterns and conditions of knowledge flows in innovation system is highly relevant for securing a sustainable business environment (Díez-Vial and Montoro-Sánchez, 2016; Sorenson et al., 2006). Especially, the role of inter-firm collaboration as a mechanism for sharing technological knowledge and the importance of knowledge transferability/multi-connectivity in promoting efficient R&D investment have been intensively addressed in the discussion of innovation management (Cantner and Rake, 2014; Guan and Liu, 2016; Williams and Lee, 2009). In this context, maximizing the potential benefits arising from the network interaction has been seen as a key factor that provides a source for new growth (Shin and Park, 2010).

Accordingly, a typical knowledge (flow) network can consist of a diverse set of exchange objects, such as patent co-inventorship, formal strategic alliance and employee interaction. This implies that a knowledge flow can be measured by a number of proxy variables. For scholars within the field of innovation management, technological knowledge flows are measured by patent data, which constitute a direct outcome of inventive efforts. The flow of patent citations (Park et al., 2018) and co-classification approach (Mun et al., 2019b) are well-known examples of capturing technological knowledge flows, especially in the domain of interdisciplinary research. While citation analysis uses both “backward” and “forward” citations to measure the knowledge inflow and outflow structure of technical fields, patent co-classification approach relies on quantifying the co-occurrence of classification codes to study the relatedness between technology sectors (Lim and Park, 2010). The relationship between patent citations and technological knowledge flow is analogous to citations in academic articles. Hence, a patent citation network can identify the structure and the characteristics of knowledge flow, because a citation can reveal the potential diffusion of information

from the cited to citing patent (Chen, 2017). The patent co-classification analysis refers to the investigation of co-occurring classification codes within the patent data. In presence of multiple classification codes, it is assumed that there is a dependency between these codes, leading to knowledge flows among technology areas from a technological perspective. As each classification code is responsible for describing a specialized technological knowledge area, patterns of knowledge flows can become apparent when looking at their interrelationships (Song et al., 2017). Additionally, the co-classification analysis has also found application in the construction of M&A interaction network for visualizing the dynamics of interacting markets (Aaldering et al., 2019a).

Within the current workflow, we adopted the co-classification approach to explore the degree of interactions between the segments pertaining to the process industries and the segments representative for digital transformation. Although data collected from Crunchbase do not directly provide information about knowledge flows, knowledge interactions can be generated by measuring the co-occurrence of industry category codes. As briefly mentioned in the previous subsection, every new startup registered at Crunchbase receives at least one or multiple category codes that are most suitable for describing the industry in which the startup operates. Hence, each category code represents an industry and is comparable to patent classification codes in case of patent co-classification approach. In this respect, the relation between two co-occurring industry category code can serve as a proxy for measuring the interacting industries. To achieve this, a knowledge interaction matrix (e.g. co-occurrence matrix) was firstly created using the industry category codes. Subsequently, the resulting matrix was transformed into a network representation to facilitate the intuitive understanding of interaction dynamics. This is a suitable way of visualizing interdisciplinary knowledge interconnectedness. Subsequently, the quantified interaction dynamics were used as a starting point for formulating further indices that help answer the research questions of interest. Fig. 1 provides an exemplary illustration of how the co-classification analysis is conceptualized in the current study.

Using the industry category codes, the co-occurrence matrix can be constructed by aggregating the number of linkages between two interacting industries. This matrix has a symmetric $n \times n$ structure, whereby the numbers represent the frequency of interactions. Since there is no directional dependency between category codes, the relations are undirected. Consequently, the nodes indicate the industry category codes, whereby the links are indicative of the intensity of relations between these codes. The thickness of the links is proportional to the intensity of interactions.

3.3. Exploration of digital influence on process industries

To further explore the extent of digital transformation on the process industries, additional analyses were undertaken. In the context of this study, the so-called IT-affinity refers to the degree to which selected industries embrace and implement new digital technologies. Hence, it represents a positive attitude towards usage behavior of digital and AI-enabled solutions based on cognitive instrumental processes (Venkatesh and Davis, 2000; Yang and Shih, 2020). Accordingly, we applied a structured approach that focuses only on the illustration of interactions between the segments pertaining to the process industries and the industries representative for digital transformation. The concept of this approach is depicted in Fig. 2. Herein, only a snippet of the whole interactions is shown due to its conceptual character.

Similar to the construction of co-occurrence matrix in the previous subsection, the first step is the transformation of the specific IT-interactions with each segment of the process industries into a comprehensive vector. This vector encompasses the existing and non-existing interactions of a process industry with IT-related industries. By accumulating all vectors, a comprehensive matrix can be constructed that contains all IT-related interaction. The x-axis of the matrix represents the industry segments pertaining to the cluster of process industries, whereas the heading of the matrix indicates the segments representative for digital transformation. Accordingly, Industry A and B could be one of the following: “Biotechnology”, “Chemical”, “Pharmaceutical”, “Food and Beverage”, “Energy”, “Metal and Mining”, “Oil and Gas”, “Paper Manufacturing” and “Wood Processing”. However, due to its vast and sparse nature, the resulting matrix is not intuitively interpretable. Thus, the matrix is visualized as a heat map plot. In this representation, the color of each box reflects the interaction intensity. The darker the coloration of a box, the higher the interaction intensity between interacting industries. This approach not only offers an easily understandable overview on the overall digitalization landscape of the process industries, but also highlights the dominant interaction partners from the IT sector. As such, important drivers of the digitalization and unique preferences in the choice of IT-related industries become identifiable.

Next, to investigate the uniqueness of the IT-interaction portfolio within the process industries and to quantify their differences, the Ruzicka similarity coefficient was applied (Cha, 2007). This coefficient quantifies the intersection between the IT-related industries while taking the specific interaction intensity into account. In simple terms, this coefficient uses the ratio of the intersecting set to the union set as the measure of similarity and returns a numerical value, which is representative of how closely two different areas are related. In the equation (shown in Fig. 2), A and B represent two different interaction vectors and

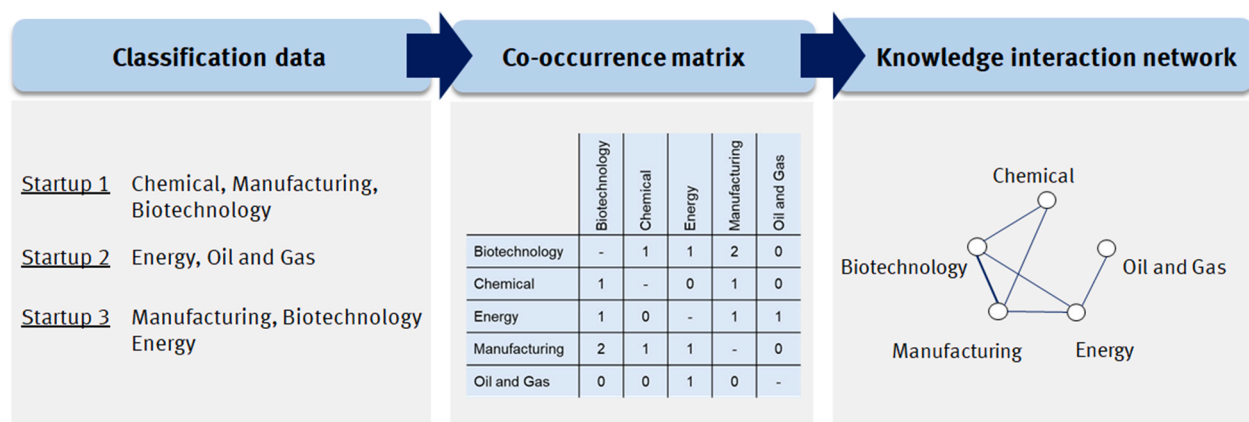


Fig. 1. Conceptual process of generating knowledge interaction network. (This is a sequential process involving the computation of co-occurrence matrix and the generation of knowledge interaction network.)

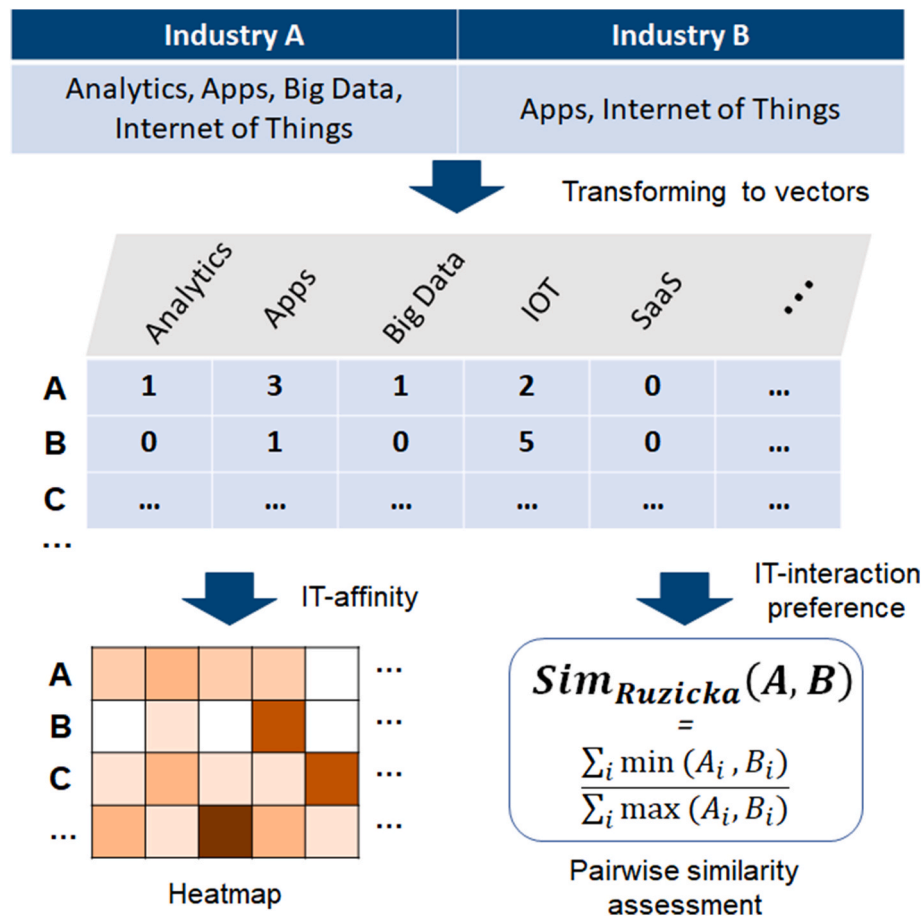


Fig. 2. Conceptual process of exploring the IT-affinity and IT-interaction preference. (Only IT-related interactions are selected to compute a weighted vector, which is subsequently visualized as a heat map plot and is given as an input for the similarity measurement.)

i stands for the index position within the vectors, where each position corresponds to the interaction rate with one of the unique IT-related industries. The results are visualized as a similarity matrix to make the results easily accessible.

4. Empirical findings

4.1. Descriptive statistics

In this subsection, selected descriptive statistics were extracted to gain an initial sense of the data. Fig. 3 describes the number of newly created startups sorted by the year of foundation. In total, 16,429 startups were founded in the period between 2010 and 2018.

Overall, the number of newly founded startups has increased

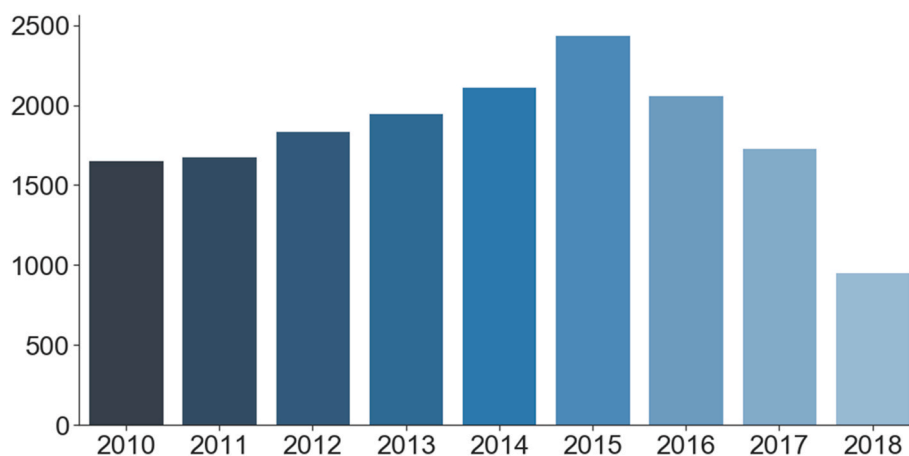


Fig. 3. Historical development trend of startups within the process industries. (The x-axis represents the founding year of startups, while the y-axis corresponds to number of startups.)

relatively steady up to 2015. However, their number began to decrease continuously from 2016 onwards. This shows a similar pattern to the development of patent applications. In general, we can observe a decline in the number of patent applications in most recent years, although we may anticipate a growth in the number of patent applications due to ever-increasing R&D investments. This can be attributed to the fact that the contents of patent applications are usually published with a delay of 18 months after the filing date of application. In the same sense, there is a certain time lag between the foundation year of a new venture and the registration year, when a startup is noticed by the global network of investors and entered into the database. Hence, the decline should not be interpreted as a shortfall in entrepreneurial activities.

Next, we turned our attention to analyzing the overall distribution of industry category codes. In this manner, we can gain a better understanding of the overall composition of the industry structure. Table 1 summarizes the top 20 frequently assigned industry category codes. There was a total of 563 different industry category codes, which are relevant for characterizing the involved markets in the process industries. The sum of frequencies is larger than the number of identified startups, because the majority of analyzed startups are characterized by more than one code. This implies that they tend to operate in an interdisciplinary business environment, where a combination of different knowledge capabilities is necessary.

In general, the relative high coverage of specific industry category codes, such as “Food and Beverage”, “Biotechnology”, “Energy”, “Pharmaceutical” and “Oil and Gas”, is a natural result of the search strategy used. As the data set is generated based on the definition of process industries (cf. Section 3.1), the corresponding industry category codes are well represented across the data. In Table 1, they are highlighted with a bluish background color. It is noticeable that the industry category codes related to pharmaceutical and biotechnology industry, such as “Health Care”, “Medical”, “Therapeutics”, “Life Science” and “Medical Device”, are also at the top of frequently assigned codes. In fact, the biotechnology industry is considered as an import pillar for the development of new therapeutic agents and clinical diagnostic testing services. Advances made in the molecular biology and genetic engineering have led to a profound transformation in the pharmaceutical and overall healthcare industry. They induced new types of research techniques and created a new organizational form made up of networks of venture capitalists, university spin-offs and large pharmaceutical firms. Especially, biopharma has emerged as a subset of the pharmaceutical industry, which manufactures pharmaceuticals that are distinct

from chemical-based drug and are biological in nature. Another important aspect is the extraordinarily high presence of “Food and Beverage”. The food industry has evolved by introducing Agri-Food tech startups, which encompass novel food products and ingredients such as alternative meat and dairy. However, due to the broad definition of “Food and Beverage”, some segments, which are not really representative for the process industries, can be observed. This point will be further explained in the next subsection. Moreover, industry category codes related to digital transformation, such as “Internet”, “Information Technology”, and “Software”, seem to play an important role in sustaining competitive advantage in the process industries. This is in line with the increasing digitalization of manufacturing processes that make existing production facilities more intelligent and contribute to co-creation of values with customers. Digital innovations have changed the way of how firms generate values and improve customer experience by facilitating end-to-end connectivity of business processes and by delivering digital channels for marketing. For example, production plants can use advanced analytics using machine learning algorithms for adjusting process parameters to optimize the production network. Moreover, internet and e-commerce are an ideal path to make the supply chain of food and beverage products more efficient. The industry category code “Manufacturing”, which is closely related to the business operations in the process industries, is moderately represented. Lastly, the high proportion of “Renewable Energy” reflects a positive association with the “Energy”.

4.2. Investigation of industry interaction dynamics

Following the conceptual process described in subsection 3.2, we constructed the knowledge interaction network to develop insights into interacting industries. In Table 2, we illustrated the top 20 frequently interacting pairs of industries. This relation is extracted from the interaction network through quantification and can highlight whether there are tendencies for exchange of knowledge within or outside of the conventional industrial boundaries. The frequency with which two different category codes appear together indicates the amount of interaction dynamics between distinct industries. In total, 14,542 different interacting pairs were present in the data set.

According to above table, no remarkable interactions between two distinct industries were identified. This implies that there are no unexpected patterns in the distribution of pairwise interactions. For example, the most frequently occurring interactions, such as “Biotechnology–Health Care”, “Biotechnology–Medical” or

Table 1

The top 20 frequently assigned industry category codes. (The codes belonging to the process industries are marked with a bluish background color.)

Industry category codes	Number of frequencies
Food and Beverage	6066
Biotechnology	5753
Health Care	3159
Energy	2574
Pharmaceutical	2070
Medical	1607
E-Commerce	1524
Manufacturing	1104
Renewable Energy	957
Oil and Gas	853
Information Technology	851
Therapeutics	762
Life Science	753
Agriculture	748
Internet	713
Software	706
Food Processing	638
Medical Device	638
Restaurants	529
Solar	513

Table 2

Interacting pairs of industry category codes. (Only the top 20 frequently occurred interaction pairs are listed.)

Industry category codes	Number of frequencies
Biotechnology – Health Care	2114
E-Commerce – Food and Beverage	1393
Biotechnology – Medical	1271
Biotechnology – Pharmaceutical	980
Energy – Renewable Energy	899
Health Care – Medical	751
Health Care – Pharmaceutical	712
Biotechnology – Life Science	691
Biotechnology – Therapeutics	673
Food Processing – Food and Beverage	624
Food and Beverage – Internet	611
Food and Beverage – Health Care	610
Biotechnology – Medical Device	530
Food and Beverage – Restaurants	527
Agriculture – Food and Beverage	512
Medical – Pharmaceutical	490
Energy – Solar	484
Food Delivery – Food and Beverage	481
Food and Beverage – Hospitality	456
Food and Beverage – Retail	403

“Biotechnology–Pharmaceutical”, were rather foreseeable because of their thematic proximity to each other. The same applies to “Energy–Renewable Energy” and “Food Processing–Food and Beverage”, which share common goals and have many similar characteristics. Interactions that do not fall under the group of conventional interaction dynamics within the process industries were not immediately visible.

On the left side of Fig. 4, we can see the complete interaction network, which is too complex to be intuitively understood. To remedy this problem, we highlighted the top 5% of most frequently occurring interactions. The threshold was set to 5%, as it enabled a clear overview of significant interactions. These are interactions that deserve more attention and are shown on the right in Fig. 4. Interestingly, the depicted interactions form a singular coherent network rather than isolated networks. This observation indicates that the process industries draw knowledge from similar adjacent industries. Herein, blue nodes indicate segments that pertain to the family of process industries, while gray nodes denote all other remaining industries. The size of the nodes correlates with the absolute number of connections assigned to each node, whereby the thickness of edges shows the degree of interaction intensity. Upon inspecting the highlighted network, it becomes evident that a substantial part of interaction dynamics is related to the digitalization of business processes. If we pay attention to how the nodes are connected with each other, we observe that the IT-related industry category codes are positioned at the center of the highlighted network. For example, industries such as “Software”, “Artificial Intelligence”, “Information Technology” show not only good connectivity with various industries belonging to the family of process industries but also are well interconnected with each other. The relatively large size of these nodes confirms the fact that these industries play an important role within the process industrial landscape. With regard to utilizing the opportunities arising from digital transformation, some industries are relatively less affected, as automatization and digitalization levels are already high or the potential to add digital contents to their products is limited. However, the use of advanced digital solutions, such as “Machine Learning”, “Artificial Intelligence”, “Internet of Things” and “Big Data”, is expected to disseminate more gradually, as massive data stream generated through the sensors and connected devices require the benefits of improved simulation and modeling capabilities. For example, “Internet of Things” represents a network of physical objects that are digitally connected to sense, monitor and interact among participants in the supply chain to facilitate timely planning, control and coordination of the supply chain processes. Hence, it provides a common operating picture for innovative applications. In contrast, “Big Data” refers to an

umbrella term for any technique to process a large amount of data, including capture, transfer, storage, curation, search, analysis, visualization, security, and privacy. There is an enormous amount of data generated from the whole manufacturing chain consisting of raw material supply, manufacturing, product distribution, logistics and customer support. The exploitation of “Big Data” could contribute to improved factory operations, reducing machine downtime and enhancing supply chain efficiency.

Furthermore, value chain partners wish to have a seamless digital experience, even from the companies in the process industries. Especially, in the field of pharmaceuticals, advances in machine learning approaches not only help improve supply chain efficiency by streamlining the manufacturing processes but also expedite the drug discovery process by enabling large-scale compound libraries to be exploited (Chen et al., 2018). “E-commerce” seems to mainly influence the food and beverage industry. As already mentioned in subsection 4.1, the broad definition of “Food and Beverage” has led to the conclusion that the process industries may include segments which do not correspond exactly to the proposed definition in all aspects. Hence, we observe interactions pairs, such as “Food Delivery – Food and Beverage”, “E-Commerce – Food and Beverage”, and “Food and Beverage – Restaurants”. Nevertheless, this should not be interpreted as an inappropriate representation of process industries. This issue is present in all kinds of taxonomies or typologies and points to the concept of “family resemblance”, which demonstrates the lack of clear boundaries that characterize different uses of the same concept (Yeung et al., 2012). The findings so far suggest that the analysis of co-classified industry category codes supports our basic theoretical perspective on knowledge diffusion and is a useful approach for investigating the potential of digital transformation in the process industries. Furthermore, as stated in connection with Table 2, industries, which have a high degree of overlapping characteristics (e.g. “Biotechnology – Health Care” and “Energy – Renewable Energy”), are positioned close to each other, forming a kind of sub-cluster within the highlighted network.

In Fig. 5, we illustrated the interacting pairs of industries that consistently increased or decreased their intensity of interactions over time. To this end, we divided the evaluation period into three different time intervals (2010–2012; 2013–2015; 2016–2018) to capture the evolutionary dynamics. Afterwards, the associated interactions are marked with colored links (red: decreasing, green: increasing).

On the left side of above figure, we highlighted the interacting industry pairs that continuously decreased the frequency of their interactions. There were 946 decreasing interaction pairs. In contrast,

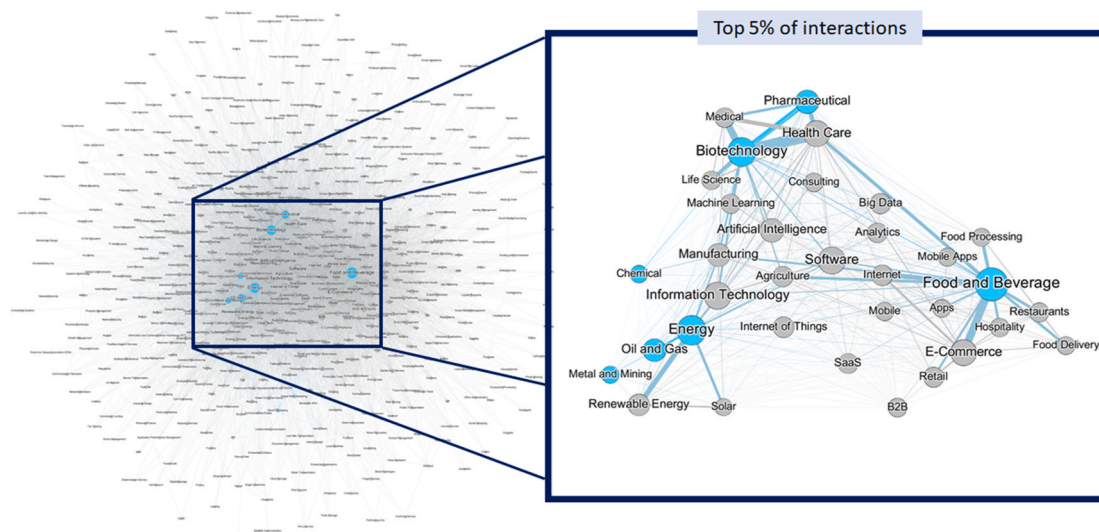


Fig. 4. Illustration of knowledge interaction network (Top 5% of interactions are highlighted).

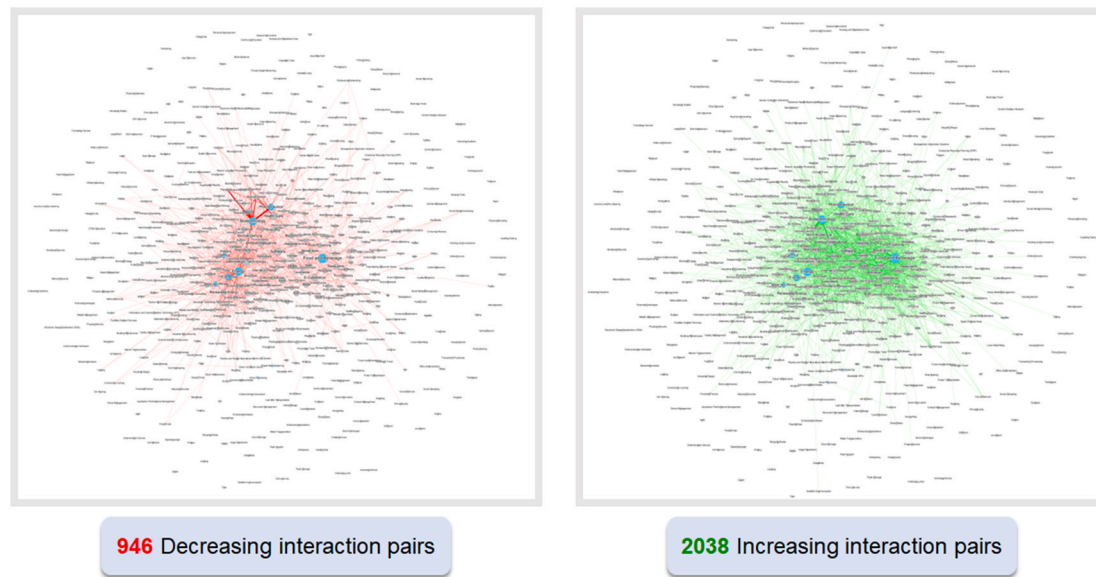


Fig. 5. Illustration of continuously decreasing and increasing interaction pairs of industries. (Left: Decreasing interaction tendencies are marked with red links; Right: Increasing interaction tendencies are marked with green links.). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

there were 2038 interaction pairs that showed increasing interaction tendencies. However, the decreasing tendency should not necessarily convey the impression that these industries will lose their relevance in the future time interval. Instead, this could be interpreted as an indication of shifting focus within the startup landscape. Fig. 6 summarizes some of the most trending interactions, which involve at least one segment from the family of process industries and demonstrate significant increase or decrease in the intensity of interactions over the years.

A marked increase in the number of continuous interactions occurred predominantly between interacting pairs related to digital transformation and the process industries, while the decline was mainly characterized by interacting pairs pertaining to the process industries and their associated disciplines. For example, “Food and Beverage” substantially increased its interaction with “Artificial Intelligence” and “Big Data” by 860 and 567 percent from the first interval to the last interval. Moreover, “Machine Learning” served as an important driver in accelerating the digitalization of “Energy”, “Oil and Gas”, “Biotechnology” and “Pharmaceutical” industry. Indeed, energy startups can embrace machine learning and its possible applications to model the energy demand and deliver critical information needed to optimize the

operational costs of renewable power generation. Especially, the huge inflow of data, collected via connected sensors placed along pipelines in oil and gas fields, can be processed and analyzed to improve the cost structure of energy companies. Moreover, in the field of pharmaceutical and biotechnology production and quality management, machine learning approaches provide a set of tools that can optimize the development and manufacturing processes. The resulting parameter estimates are capable of assessing whether and which product development models can be scaled up to manufacturing models. On the other hand, the continuously decreasing interactions occurred mostly between industries from the family of process industries. Herein, it can be argued that the classical relations linked to each other in a value-adding production chain become less important. The interactions between “Manufacturing” and “Chemical” as well as “Manufacturing” and “Metal and Mining” are prime examples of such situations. Surprisingly, the intensity of interactions between “Pharmaceutical” and “Biotechnology” showed a decreasing trend. This points to the fact that the manufacturing barriers to biologics competition and innovation have been lowered over decades and biologics-based biotechnology companies can expand their competence uncoupled from the pipeline

Interacting market segments			Change
Exemplary excerpt	Natural Resources	Oil and Gas	- 81 %
	Natural Resources	Energy	- 79 %
	Manufacturing	Chemical	- 77 %
	Manufacturing	Metal and Mining	- 70 %
	Consulting	Energy	- 64 %
	Manufacturing	Energy	- 59 %
	Oil and Gas	Energy	- 58 %
	Energy Efficiency	Energy	- 57 %
	Consulting	Pharmaceutical	- 53 %
	Biotechnology	Pharmaceutical	- 38 %

Interacting market segments			Change
Exemplary excerpt	Artificial Intelligence	Food and Beverage	+ 860 %
	Mobile Apps	Food and Beverage	+ 605 %
	Big Data	Food and Beverage	+ 567 %
	Blockchain	Energy	+ 567 %
	Machine Learning	Oil and Gas	+ 433 %
	Machine Learning	Food and Beverage	+ 350 %
	Machine Learning	Biotechnology	+ 250 %
	Artificial Intelligence	Energy	+ 250 %
	Machine Learning	Energy	+ 175 %
	Machine Learning	Pharmaceutical	+ 150 %

Fig. 6. Exemplary excerpt from decreasing and increasing interactions over time. (Left: Continuously decreasing interaction pairs involving at least one industry from the family of process industries are shown; Right: Continuously increasing interaction pairs involving at least one industry from the family of process industries are shown.)

development of pharmaceutical industry (Kinch, 2015). Furthermore, it can be noted that the decreasing importance of “Consulting” in case of “Energy” and “Pharmaceutical” correlates positively with advances in the application of machine learning techniques, implying that industry managers put more trust into data-driven operations instead of expert opinions.

Overall, such discrepancy arises from the fact that the growing need for digital agility has opened new ways for startups to deal with data governance and predictive analytics as well as to establish digitally enabled businesses that are less capital-intensive and do not depend on physical environment. This contrasts with segments operating in the process industries that are characterized by complex value chains, high asset intensity and strong anchoring to physical locations (Lager et al., 2013). The notable increase in the described interaction dynamics also reflects the opportunity seeking behavior of startups, who take advantage of the low barrier to entry that IT technologies provide. Insights gained from this exploratory analysis further suggest that emerging startups have recognized the opportunities offered by digitalization, while digitalization is not necessarily part of the business strategy of incumbent companies, struggling to realize the full potential of transformative character of digitalization.

4.3. Comparison of IT-affinity in the process industries

In this subsection, we analyzed whether certain segments within the family of the process industries demonstrate a higher degree of affinity towards digital solutions. As already communicated, the pursuit of digital leadership can entail several benefits for organizations within the process industries, which generally rely on complex asset intensive production systems. The benefits include, but are not limited to, optimization of supply chain planning, improvement of the total production process and evaluation of business models through addition of digital value-added services. Thus, having a solid knowledge on IT-affinity could provide valuable information for R&D managers and policy makers seeking to chart trends for the coming years.

In the examined data set, there were 112 different unique IT-related industries which underwent an interaction with a segment from the process industries. In Fig. 7, a heat map visualization is used to shed light on whether varying inclinations towards IT-related industries are existent.

On the heat map, only interactions with 25 different IT-related industries were depicted, as these account for over 85 percent of all existing interactions. The colors on the heat map correspond to the level of interaction intensity between interacting industries. The darker the color, the more frequent interactions occur between industries, and thus the greater likelihood of knowledge spillovers between them. Therefore, the IT-affinity is higher at spots where the color spectrum is more dominant. In this context, it can be observed that the more general IT-related industries, such as “Information Technology”, “Analytics” and “Software”, are among the strongest interaction partners of most process industries. Their high interaction intensity highlights, to a certain extent, the primary importance of a basic IT-infrastructure in the process industries. Apart from these fundamental IT-industries, the heat map also reveals a high level of interaction intensity between certain segments of process industries and more modern digitally-led industries, such as “Big Data”, “Artificial Intelligence” and “Machine Learning”. These industries stem from the intersection of computer science and statistics and represent one of the most rapidly growing technical fields (Jordan and Mitchell, 2015). Due to their application focus, they exceed the requirements of a basic IT-infrastructure and emphasize the movement towards a more advanced and data-driven decision making, elevating the degree of automation and digitalization to the next level. Consequently, these industries lie at the core of the realization of industry 4.0.

With that in mind, there are two more striking observations to be made in this context. First, “Biotechnology”, “Pharmaceutical”, “Food and Beverage”, “Energy” and “Oil and Gas” are the main industries that showed significantly higher affinity towards digital solutions. The more conservative segments among the process industries (e.g. “Wood Processing”, “Paper Manufacturing” and “Chemical”) showed a slower adoption rate with regard to digital capabilities. This is partly related to the fact that there are only a low number of startups concentrating on these industries. From a startup perspective, this could serve as an indication that these industries have a substantial innovation and digitalization deficiency. Secondly, some clear dissimilarities in the preferences concerning the types of chosen digital capabilities are evident among the process industries. For example, whereas both “Biotechnology” and “Pharmaceutical” share strong interactions to “Artificial Intelligence” and “Big Data”, “Mobile Apps” represents a unique preference of “Food and Beverage”. Moreover, “Biotechnology”

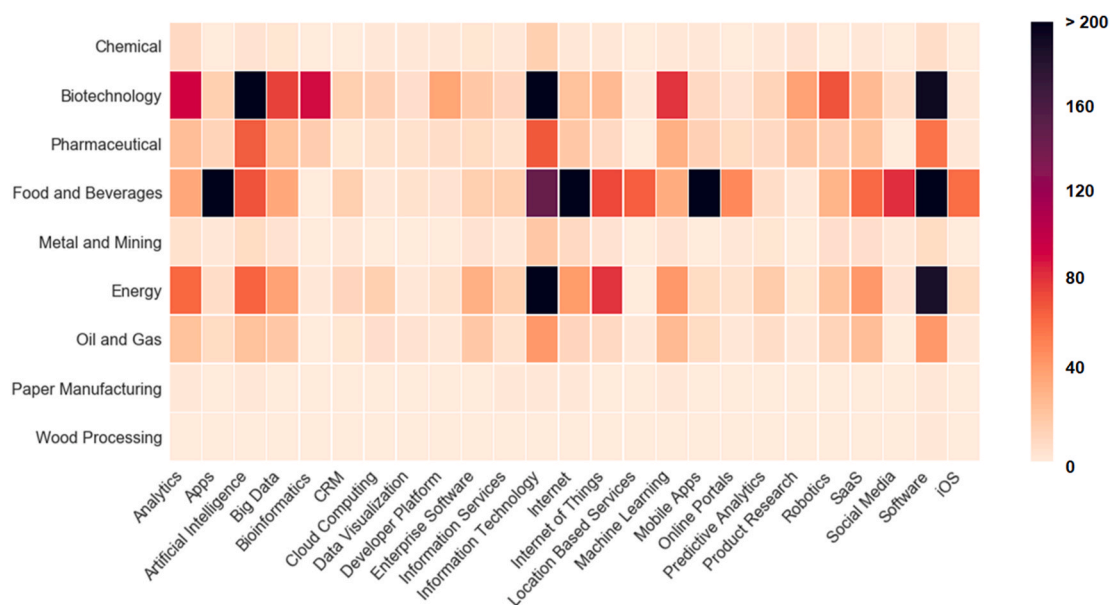


Fig. 7. Heat map visualization of IT-affinity. (The areas with stronger color intensities are those showing a higher IT-affinity.). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

demonstrates a high affinity towards “Robotics” and “Bioinformatics” - a preference only weakly observable in other segments of process industries. This indicates that digitalization does not move in a singular direction for all process industries but rather that each sector of the process industries requires a unique pathway towards industry 4.0. Overall, given the excitement about the adoption of digital opportunities, we could observe that there are still opportunities to upgrade the IT-capabilities for specific process industries. In this regard, the heat map representation could serve as a kind of morphological box to identify vacant digital opportunities within the process industries. For example, there were only a low level of interaction with industries such as “Cloud Computing” and “Data Visualization”.

Building on the illustrated differences between the process industries, we further investigated the extent of the unique IT-interaction preferences from each sector of the process industries. For this purpose, we calculated the similarity score among the segments of process industries based on the pairwise similarity measurement. This approach allows us to take both the intersection of their interactions with IT-related industries and the associated interaction intensity into account. The results are visualized in form of a similarity matrix shown in Fig. 8. This matrix depicts the similarity in IT-related interactions between each segment of the process industries in both color and similarity value. Herein, the darker the color and the higher the similarity value, the more similar are the two segments of process industries with regard to their IT-interaction preference. Accordingly, “Pharmaceutical” industry behaves similar to “Oil and Gas”, while “Energy” and “Biotechnology” are characterized by a similar set of IT-portfolio.

While some combinations show significant similarity scores (e.g. “Metal and Mining” and “Chemical” or “Biotechnology” and “Energy”), the most noteworthy similarity score is observable between “Oil and Gas” and “Pharmaceutical”. A possible explanation might be found in the heat map visualization of the interactions of the process industries with IT-related industries in Fig. 6. Apart from “Information Technology” and “Software”, “Oil and Gas” and “Pharmaceutical” show high interaction rates with “Artificial Intelligence” and “Machine Learning”. This might be related to the dependence of segments of the process industries on sophisticated production systems, which offer plentiful application possibilities for novel digital technologies. Yet, whether this similarity of preference in digital capabilities also leads to similar fields

of practical applications requires further research.

The illustrated similarity matrix revealed that the similarity scores between the preferences in IT-related capabilities are quite low between most of the process industries. This strengthens the assumption that each segment of the process industries has unique requirements in terms of digital technologies and thus needs to adopt different approaches in the realization of industry 4.0.

5. Discussion

As digitalization is not an isolated phenomenon that occurs only in few companies, it is important not to restrict the analysis of trends to the level of individual companies, but to take a wider view of the whole sector. Our research approach, therefore, differs from other studies in that it uses a broad definition of the process industries. The data set includes not only companies that can be classified as process industries in the strict sense. It also includes companies that are located above and below these companies in the value chain, thus giving a comprehensive picture of the whole business ecosystem. Based on this perspective, our study revealed the IT-preferences of an entire sector and showed the progressive digital change to which typical process companies have to adapt in these sectors.

5.1. Using startup information as a means for investigating the influence of digital technologies on the process industries

Regarding the first research question, our empirical findings demonstrated that startup information can provide valuable insights into interacting pairs of industries, thereby quantifying the degree of distribution and variation of digital technologies within the process industries. Industry category codes within the startup data represent industries (e.g. knowledge areas), whose assessment may be useful for revealing new application areas of digital technologies. Subsequently, we adopted the concept of knowledge flow to generate an interaction network and visualized the varying degree of digital adoption within the process industries. We were able to show that service-intensive and digitally-driven sectors make up a substantial part of the innovation ecosystem of process industries. The findings were able to provide a new perspective at the level of market interactions that were not able to

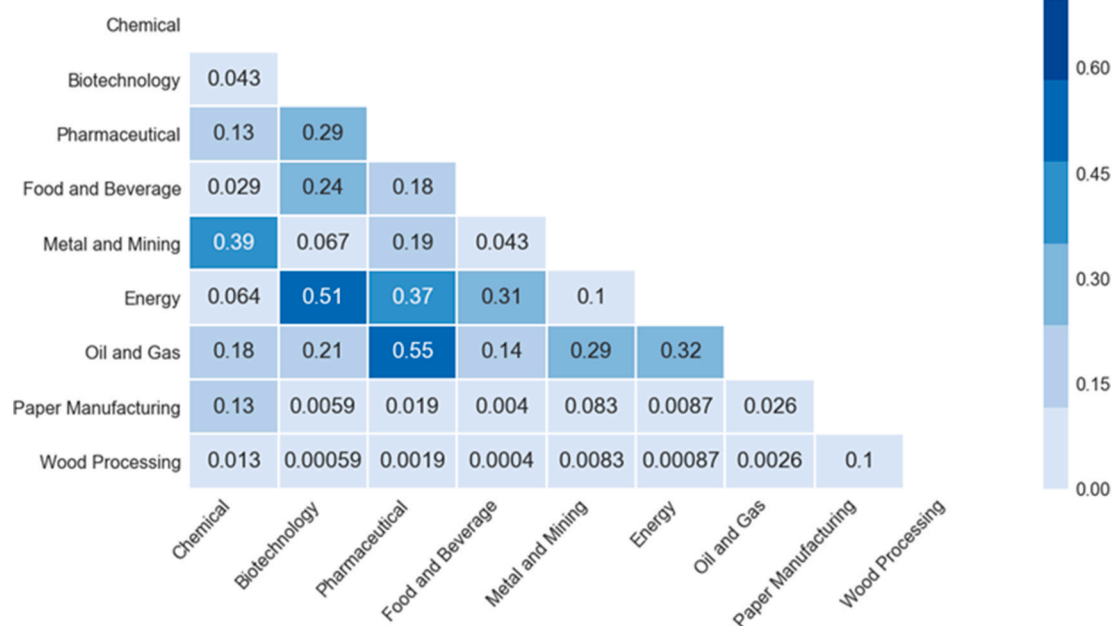


Fig. 8. Matrix visualization of IT-interaction preference similarity. (The areas with stronger color intensities are those showing a higher similarity in terms of IT-portfolio.). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

portray using the patent information (Aaldering et al., 2019a). Thus, startup data are an appropriate means for scouting new market trends beyond the mere technological dimension (Marra et al., 2020).

5.2. Industries that act as drivers for digital transformation

Coming back to our second research question, the findings did provide evidence for the driving forces behind the digital transformation of process industries. By analyzing the continuously increasing interactions involving at least one segment from the process industries, we observed that “Machine Learning” and “Artificial Intelligence” played a leading role in accelerating the digitalization of process industries. “Mobile Apps” also significantly increased its interaction with “Food and Beverage”, thereby helping the industries move closer to a consumer-oriented business model. This could be related to bringing the ordering process into the digital age, and booming of the e-commerce food and grocery delivery market (Kapoor and Vij, 2018). Furthermore, “Block Chain” acted as a driving force behind the digitalization of “Energy”. With energy systems undergoing substantial changes to accommodating the increasing volumes of embedded renewable generation, block chain technology can help establish real-time data management and move carbon credits or renewable energy certificates onto the block chain for secure transactions (Andoni et al., 2019).

5.3. Distinguishing the affinity with regard to adopting digital technologies

Regarding the last research question, a remarkable difference in the affinity of the various digital technologies was observed. While “Biotechnology”, “Food and Beverage”, “Energy” and “Oil and Gas” showed a similarly high degree of IT-affinity, “Chemical”, “Metal and Mining”, “Paper Manufacturing” and “Wood Processing” were rather shy in adjusting their production process using digital capabilities to optimize manufacturing efficiency. In order to become competitive in long term, the latter sectors must expand their spectrum of value creation by leveraging digital technologies (María-Luz et al., 2019). In this manner, these sectors might recognize development potentials that go beyond being cost-competitive commodity producers. Hence, the digitalization process needs to be seen through business-strategic glasses to collect relevant business cases. Moreover, “Food and Beverage” showed unique affinity towards “Location Based Services”, “Online Portals” and “Mobile Apps”. This indicates that value added digital services are one of the main fields of interest in terms of creating synergies (Tao et al., 2018). Innovation marketing needs to build on collective intelligence for capturing additional growth opportunities and for predicting customer behavior from a strategic point of view.

6. Managerial and theoretical implications

The way towards industry 4.0 is complex and involves the interplay of multiple interacting process instances, raising significant challenges to managers attempting to modernize their businesses. In this context, we presented a data-driven exploratory research approach that sheds light onto the ongoing digital transformation within the various segments of the process industries. Overall, this paper contains following managerial and theoretical implications. With regard to managerial contributions, we discovered multiple aspects that could serve both R&D managers and policy makers as an important guideline in their current digitalization efforts.

1 Foster cross-sectoral learning

Our findings have shown that there is no universally valid path towards industry 4.0, as each segment within the family of the process industries has demonstrated unique preferences with regard to IT-related capabilities. However, we have also shown that there is significant overlap between specific segment combinations. While the

differences could be harnessed to identify suitable new digitalization avenues, the similarities might help to streamline the implementation of capabilities that are already established in other segments of the process industries by profiting from their lessons learned and identifying suitable complementary digital capabilities. Hence, the digitalization behavior of leaders can be copied by laggards to set strategic goals and create an incentive plan. Consequently, learning from the ongoing digitalization efforts in neighboring areas could provide substantial benefits for one's own digitalization projects.

2 Vacant digitalization opportunities

From a more application-oriented perspective, our approach was not only able to reveal emerging digital trends within the process industries, but it could also empower the identification of vacant digital and entrepreneurial opportunities. The utilized heat map visualization of the IT-interaction preferences clearly revealed the underserved digital opportunities that could provide future growth fields for novel and incumbent companies. Therefore, our study highlights the importance of monitoring the digitalization process across all areas of process industries, as it could serve as an important new aspect in new business development.

3 Promote digital education

The role of business translators, who can link digital capabilities with domain-specific business knowledge, might become more significant to exploit vacant niches. Hence, there will be more needs for disseminating digital education, which advances the knowledge of traditional managers. Moreover, the digital education can be facilitated by creating platforms for knowledge exchange and initiating partnerships across the sectors.

With regard to theoretical contributions, our study showed the applicability of startup information for the investigation of knowledge flows and industry interaction dynamics. Thus, our study is one of the first of its kind to introduce startup information as a suitable data source for research in these fields, complementing previously used data sources, such as patents, publications and M&A-deals (Aaldering et al., 2019b; Mun et al., 2019a). In this connection, we introduced an analysis framework, which complements the survey- or interview-based studies by incorporating instruments from the data analytics toolbox (Onufrey and Bergek, 2020). Thus, our study provides other researchers a novel perspective to investigate the digital transformation beyond merely considering business models (Bouwman et al., 2019). It provides scholars (especially from the social science disciplines) with new stimuli to investigate industry phenomena. Furthermore, this paper is one of the earliest attempts to combine research domains of entrepreneurship, process industries and the context of innovation management. Thus, this research paved the way for subsequent researches aiming at illuminating the intersection of these disciplines (Marra et al., 2020). Lastly, this paper responds to the call for broadening the range of hypotheses and instruments for the field of innovation and entrepreneurship (Douglas et al., 2020; Shepherd, 2015).

7. Conclusions and outlook on future research

The volatile industry environment requires companies across all industries to continuously innovate and familiarize with changing customer requirements. This also applies for the process industries, which traditionally relied on producing commodities at a low cost. One way to promote innovation within the process industries is by adopting digital capabilities. In this context, startups are considered as a strong driver of digital transformation due to their high flexibility and explorative opportunity seeking behavior. While startups as a research subject have received increased attention from scholars in the field of innovation management, their role in the digital transformation of

process industries has been overlooked. More specifically, the influence of digitalization at the level of process industries has not received its deserved attention in scholarly research. Subsequently, quantification and visualization of interconnectedness between digital and process-based industry segments from a startup perspective can be an important step towards understanding the digital evolution of industries.

Based on the data-driven analysis, we demonstrated that startup data are a suitable source for the investigation of digital transformation. Unlike the expectations, not all segments of process industries are laggards in regard to incorporating digital capabilities. Especially, industries such as “Biotechnology”, “Pharmaceutical”, “Food and Beverage”, “Energy” and “Oil and Gas” demonstrated a stronger IT-interaction preference, rendering them to leaders in the race towards industry 4.0 compared to their peers within the process industries. However, we also observed that there is no universal strategy in the adoption of digital capabilities. Each industry struck a different path depending on their specific conditions and requirements. In sum, this study contributes to debates in the academic community on the issue of whether individual segments of the process industries behave differently when adopting digital capabilities and what opportunities arise for them. This study uncovered vacant digitalization opportunities that could be exploited to strengthen the integration process of digital solutions into existing industry infrastructure. Moreover, this study provides a new analysis framework that is intuitive to understand the industry interactions and provides new stimuli to investigate industry phenomena by employing techniques from the modern data analytics toolbox.

As with all research, the current study is subject to several limitations. First, as our study focuses on startups, it can only reveal and explore digital trends and capabilities that are accelerated by new ventures. Digital trends and capabilities introduced by incumbent firms remain undiscovered. Accordingly, this study shows only a snapshot of the ongoing digitalization in the process industries. To remedy this, future research could combine multiple data sources (e.g. patents or M&A) to gain a more comprehensive picture of this phenomenon.

Appendix

Appendix A. Glossary of technical terms - Adapted from (Ben-Daya et al., 2019; Büchi et al., 2020; Alejandro Germán Frank et al., 2019; Gürdür et al., 2019; L'Heureux et al., 2017; Plessis and Boon, 2004; Wang et al., 2016)

Technical terms	Definition
Digitalization	Enabling, improving, and transforming business operations, functions, models or activities by leveraging digital technologies.
Analytics	Technologies and processes of acquiring in-depth knowledge and understanding from the data.
Apps	Types of software applications accessed via the internet browser and specifically designed for fluidly responding to contents viewed on a smartphone.
Artificial Intelligence (AI)	Ability of a computer or a computer-enabled robotic system to process information and produce outcomes in a similar manner associated with human intelligence.
Big Data	An umbrella term for any technique to process a large amount of data, including capture, transfer, storage, curation, search, analysis, visualization, security, and privacy.
Customer relationship management (CRM)	Building and managing of customer relationships on an organizational level through understanding, anticipating and managing of customer needs, based on knowledge gained of the customer.
Cloud Computing	Enabling on-demand network access to a shared pool of computing resources, thereby facilitating the archiving and processing of large quantities of data with high performance in terms of speed, flexibility, and efficiency. It results in a greater number of services developed based on data for a productive system – including monitoring and control functions – to ensure quality and improve operations and production.
Data Visualization	Tool in data analytics, which enables to visually detect complex structures and patterns in the data.
Industry 4.0	Transformation from machine dominant manufacturing to digital manufacturing and bears a huge potential to achieve a higher level of personalization and resource efficiency.
Information Services	Services available via internet for providing information
Information Technology	Technologies involved in information processing and handling, such as computer hardware, software, telecommunications.
Internet of Things (IOT)	Network of physical objects that are digitally connected to sense, monitor and interact among participants in the supply chain to facilitate timely planning, control and coordination of the supply chain processes.
Location Based Services	Services offered through mobile devices by taking into account the geographical location of the devices. It can be used as a tracking of non-human object for logistical purposes.
Machine Learning	Ability of a computer system to learn and recognize patterns from data with the aim of providing data-driven insights, decisions, and predictions.
Mobile Apps	Types of applications specifically designed for mobile devices and are built for a specific platform, such as iOS and Android.

(continued on next page)

Second, in terms of data used, our study relied solely on a single database. This increases the likelihood that new ventures, which are not registered at Crunchbase, cannot be considered. Apart from that, our analysis was dependent on the classification scheme provided by Crunchbase. Therefore, its depth of analysis is limited to the granularity of Crunchbase's classification scheme. Hence, our data set might also include startups that do not perfectly fit into the definition of process industries. Moreover, industry niches that are not covered within this classification scheme are not considered within the study. In this context, future research could develop and apply a customized classification scheme that is adapted for their specific research context.

Finally, this paper offers some further interesting perspectives for future research directions. When it comes to explaining the drivers, further research can focus on investigating the motivation of adopting digital capabilities. These might help understand the technological, competitive and cultural requirements of modernizing the production processes. Another important aspect is the consideration of geographical scope of the used data, as there might be some country-specific differences in adopting digital transformation. The current study, however, did focus on mapping general pathways instead of illustrating individual country-specific pathways. Hence, future research can address this issue by comparing country-specific behaviors. Furthermore, a more in-depth analysis of the technology-focused applications of these digital capabilities could open another new research avenue, as it can provide a more granular insights into the differences of the digitalization pathways.

Declarations of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

(continued)

Technical terms	Definition
Online Portals	Websites that offer users access to a wide range of information, tools and links.
Predictive Analytics	Enabling advanced predictive capacity and identifying events that can affect production before it happens based on data analytics.
Software as a Service (SaaS)	A cloud-based (on-demand) service model that provides service consumers without installing any application locally to outsource the operation of their business applications to a SaaS provider.

References

- Aaldering, L.J., Leker, J., Song, C.H., 2019a. Uncovering the dynamics of market convergence through M&A. *Technol. Forecast. Soc. Change* 138, 95–114. <https://doi.org/10.1016/j.techfore.2018.08.012>.
- Aaldering, L.J., Leker, J., Song, C.H., 2019b. Recommending untapped M&A opportunities: a combined approach using principal component analysis and collaborative filtering. *Expert Syst. Appl.* 125, 221–232. <https://doi.org/10.1016/j.eswa.2019.02.004>.
- Aaldering, L.J., Leker, J., Song, C.H., 2018. Analyzing the impact of industry sectors on the composition of business ecosystem: a combined approach using ARM and DEMATEL. *Expert Syst. Appl.* 100, 17–29. <https://doi.org/10.1016/j.eswa.2018.01.045>.
- Abu-Saifan, S., 2012. Social entrepreneurship: definition and boundaries. *Technol. Innov. Manag. Rev.* 2, 22–27. <https://doi.org/10.22215/timreview/523>.
- Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., Peacock, A., 2019. Blockchain technology in the energy sector: a systematic review of challenges and opportunities. *Renew. Sustain. Energy Rev.* 100, 143–174. <https://doi.org/10.1016/j.rser.2018.10.014>.
- Antolín-López, R., Céspedes-Lorente, J., García-de-Frutos, N., Martínez-del-Río, J., Pérez-Valls, M., 2015. Fostering product innovation: differences between new ventures and established firms. *Technovation* 41–42, 25–37. <https://doi.org/10.1016/j.technovation.2015.02.002>.
- Argote, L., Ingram, P., 2000. Knowledge transfer: a basis for competitive advantage in firms. *Organ. Behav. Hum. Decis. Process.* 82, 150–169. <https://doi.org/10.1006/obhd.2000.2893>.
- Autio, E., Nambisan, S., Thomas, L.D.W., Wright, M., 2018. Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems. *Strateg. Entrep. J.* 12, 72–95. <https://doi.org/10.1002/sej.1266>.
- Basole, R.C., Park, H., Barnett, B.C., 2015. Coopetition and convergence in the ICT ecosystem. *Telecommun. Pol.* 39, 537–552. <https://doi.org/10.1016/j.telpol.2014.04.003>.
- Basole, R.C., Park, H., Chao, R.O., 2019. Visual analysis of venture similarity in entrepreneurial ecosystems. *IEEE Trans. Eng. Manag.* 66, 568–582. <https://doi.org/10.1109/TEM.2018.2855435>.
- Becker, J.U., Clement, M., Nöth, M., 2016. Start-ups, incumbents, and the effects of takeover competition. *J. Bus. Res.* 69, 5925–5933. <https://doi.org/10.1016/j.jbusres.2016.05.005>.
- Ben-Daya, M., Hassini, E., Bahrour, Z., 2019. Internet of things and supply chain management: a literature review. *Int. J. Prod. Res.* 57, 4719–4742. <https://doi.org/10.1080/00207543.2017.1402140>.
- Benitez, G.B., Ayala, N.F., Frank, A.G., 2020. Industry 4.0 innovation ecosystems: an evolutionary perspective on value cocreation. *Int. J. Prod. Econ.* 228, 107735. <https://doi.org/10.1016/j.ijpe.2020.107735>.
- Bouwman, H., Nikou, S., de Reuver, M., 2019. Digitalization, business models, and SMEs: how do business model innovation practices improve performance of digitalizing SMEs? *Telecommun. Pol.* 43, 101828. <https://doi.org/10.1016/j.telpol.2019.101828>.
- Breitzman, A., Thomas, P., 2015. The Emerging Clusters Model: a tool for identifying emerging technologies across multiple patent systems. *Res. Pol.* 44, 195–205. <https://doi.org/10.1016/j.respol.2014.06.006>.
- Brozzi, R., Forti, D., Rauch, E., Matt, D.T., 2020. The advantages of industry 4.0 applications for sustainability: results from a sample of manufacturing companies. *Sustainability* 12, 3647. <https://doi.org/10.3390/su12093647>.
- Büchi, G., Cugno, M., Castagnoli, R., 2020. Smart factory performance and Industry 4.0. *Technol. Forecast. Soc. Change* 150, 119790. <https://doi.org/10.1016/j.techfore.2019.119790>.
- Cantner, U., Rake, B., 2014. International research networks in pharmaceuticals: structure and dynamics. *Res. Pol.* 43, 333–348. <https://doi.org/10.1016/j.respol.2013.10.016>.
- Castelo-Branco, I., Cruz-Jesus, F., Oliveira, T., 2019. Assessing industry 4.0 readiness in manufacturing: evidence for the European union. *Comput. Ind.* 107, 22–32. <https://doi.org/10.1016/j.compind.2019.01.007>.
- Cavallo, A., Ghezzi, A., Dell'Era, C., Pellizzoni, E., 2019. Fostering digital entrepreneurship from startup to scaleup: the role of venture capital funds and angel groups. *Technol. Forecast. Soc. Change* 145, 24–35. <https://doi.org/10.1016/j.techfore.2019.04.022>.
- Caviggioli, F., 2016. Technology fusion: identification and analysis of the drivers of technology convergence using patent data. *Technovation* 55–56, 22–32. <https://doi.org/10.1016/j.technovation.2016.04.003>.
- Cha, S.-H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Model. Methods Appl. Sci.* 4, 300–307.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., 2018. The rise of deep learning in drug discovery. *Drug Discov.* 23, 1241–1250. <https://doi.org/10.1016/j.drudis.2018.01.039>. Today.
- Chen, L., 2017. Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations. *J. Informetr.* 11, 63–79. <https://doi.org/10.1016/j.joi.2016.04.018>.
- Culot, G., Orzes, G., Sartor, M., Nassimbeni, G., 2020. The future of manufacturing: a Delphi-based scenario analysis on Industry 4.0. *Technol. Forecast. Soc. Change* 157, 120092. <https://doi.org/10.1016/j.techfore.2020.120092>.
- Curran, C.-S., Leker, J., 2011. Patent indicators for monitoring convergence – examples from NFF and ICT. *Technol. Forecast. Soc. Change* 78, 256–273. <https://doi.org/10.1016/j.techfore.2010.06.021>.
- Dalle, J.-M., Den Besten, M., Menon, C., 2017. Using Crunchbase for Economic and Managerial Research (No. 2017/08), OECD Science, Technology and Industry Working Papers. <https://doi.org/10.1787/18151965>.
- Decker, R., Haltiwanger, J., Jarmin, R., Miranda, J., 2014. The role of entrepreneurship in US job creation and economic dynamism. *J. Econ. Perspect.* 28, 3–24.
- Deloitte, 2017. Supply Chain Start-Ups Are Coming of Age.
- Diez-Oliván, A., Del Ser, J., Galar, D., Sierra, B., 2019. Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0. *Inf. Fusion* 50, 92–111. <https://doi.org/10.1016/j.inffus.2018.10.005>.
- Diez-Vial, I., Montoro-Sánchez, A., 2016. How knowledge links with universities may foster innovation: the case of a science park. *Technovation* 50–51, 41–52. <https://doi.org/10.1016/j.technovation.2015.09.001>.
- Douglas, E.J., Shepherd, D.A., Prentice, C., 2020. Using fuzzy-set qualitative comparative analysis for a finer-grained understanding of entrepreneurship. *J. Bus. Ventur.* 35, 105970. <https://doi.org/10.1016/j.jbusvent.2019.105970>.
- Frank, Alejandro Germán, Dalenogare, L.S., Ayala, N.F., 2019a. Industry 4.0 technologies: implementation patterns in manufacturing companies. *Int. J. Prod. Econ.* 210, 15–26. <https://doi.org/10.1016/j.ijpe.2019.01.004>.
- Frank, Alejandro G., Mendes, G.H.S., Ayala, N.F., Ghezzi, A., 2019b. Servitization and Industry 4.0 convergence in the digital transformation of product firms: a business model innovation perspective. *Technol. Forecast. Soc. Change* 141, 341–351. <https://doi.org/10.1016/j.techfore.2019.01.014>.
- Frishammar, J., Lichtenthaler, U., Kurkkio, M., 2012. The front end in non-assembled product development: a multiple case study of mineral- and metal firms. *J. Eng. Technol. Manag.* 29, 468–488. <https://doi.org/10.1016/j.jengtecman.2012.07.001>.
- Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B., Ziv, G., 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosyst. Serv.* 38, 100958. <https://doi.org/10.1016/j.ecoser.2019.100958>.
- Guan, J., Liu, N., 2016. Exploitative and exploratory innovations in knowledge network and collaboration network: a patent analysis in the technological field of nano-energy. *Res. Pol.* 45, 97–112. <https://doi.org/10.1016/j.respol.2015.08.002>.
- Gürdüz, D., El-khoury, J., Törngren, M., 2019. Digitalizing Swedish industry: what is next? data analytics readiness assessment of Swedish industry, according to survey results. *Comput. Ind.* 105, 153–163. <https://doi.org/10.1016/j.compind.2018.12.011>.
- Hill, T., 2000. *Manufacturing Strategy: Text and Cases*, second ed. PALGRAVE, New York. <https://doi.org/10.1007/978-1-349-14018-3>.
- Huang, H.-C., Su, H.-N., 2019. The innovative fulcrums of technological interdisciplinarity: an analysis of technology fields in patents. *Technovation* 84–85, 59–70. <https://doi.org/10.1016/j.technovation.2018.12.003>.
- International Business Publications, 2009. *Europe E-Commerce Laws and Regulations Handbook: Strategic Information and Regulations*, volume 1. International Business Publications, USA, Washington DC.
- Isaksson, A.J., Harjunkoski, I., Sand, G., 2018. The impact of digitalization on the future of control and operations. *Comput. Chem. Eng.* 114, 122–129. <https://doi.org/10.1016/j.compchemeng.2017.10.037>.
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. <https://doi.org/10.1126/science.aaa8415>.
- Kapoor, A.P., Vij, M., 2018. Technology at the dinner table: ordering food online through mobile apps. *J. Retailing Consum. Serv.* 43, 342–351. <https://doi.org/10.1016/j.jretconser.2018.04.001>.
- Kinch, M.S., 2015. An overview of FDA-approved biologics medicines. *Drug Discov. Today* 20, 393–398. <https://doi.org/10.1016/j.drudis.2014.09.003>.
- Ko, E.-J., McKelvie, A., 2018. Signaling for more money: the roles of founders' human capital and investor prominence in resource acquisition across different stages of firm development. *J. Bus. Ventur.* 33, 438–454. <https://doi.org/10.1016/j.jbusvent.2018.03.001>.
- Kodama, F., 2014. MOT in transition: from technology fusion to technology-service convergence. *Technovation* 34, 505–512. <https://doi.org/10.1016/j.technovation.2013.04.001>.
- Kusiak, A., 2018. Smart manufacturing. *Int. J. Prod. Res.* 56, 508–517. <https://doi.org/10.1080/00207543.2017.1351644>.
- L'Heureux, A., Gröling, K., Elyamany, H.F., Capretz, M.A.M., 2017. Machine learning with big data: challenges and approaches. *IEEE Access* 5, 7776–7797. <https://doi.org/10.1109/ACCESS.2017.2696365>.

- Lager, T., 2017. A conceptual analysis of conditions for innovation in the process industries and a guiding framework for industry collaboration and further research. *Int. J. Technol. Learn. Innovat. Dev.* 9, 189–219. <https://doi.org/10.1504/IJTLID.2017.087403>.
- Lager, T., 2010. Managing Process Innovation: from Idea Generation to Implementation, Series on Technology Management. Imperial College Press, London. <https://doi.org/10.1142/p734>.
- Lager, T., Blanco, S., Frishammar, J., 2013. Managing R&D and innovation in the process industries. *R D Manag.* 43, 189–195. <https://doi.org/10.1111/radm.12018>.
- Lager, T., Storm, P., 2013. Application development in process firms: adding value to customer products and production systems. *R D Manag.* 43, 288–302. <https://doi.org/10.1111/radm.12013>.
- Lasi, H., Fetteke, P., Kemper, H.-G., Feld, T., Hoffmann, M., 2014. Industry 4.0. *Bus. Inf. Syst. Eng.* 6, 239–242. <https://doi.org/10.1007/s12599-014-0334-4>.
- Lee, C., Kogler, D.F., Lee, D., 2019. Capturing information on technology convergence, international collaboration, and knowledge flow from patent documents: a case of information and communication technology. *Inf. Process. Manag.* 56, 1576–1591. <https://doi.org/10.1016/j.ipm.2018.09.007>.
- Lim, H., Park, Y., 2010. Identification of technological knowledge intermediaries. *Scientometrics* 84, 543–561. <https://doi.org/10.1007/s11192-009-0133-8>.
- Lim, S., Kwon, O., Lee, D.H., 2018. Technology convergence in the Internet of Things (IoT) startup ecosystem: a network analysis. *Telematics Inf.* 35, 1887–1899. <https://doi.org/10.1016/j.tele.2018.06.002>.
- María-Luz, M.-P., José-María, S.-L., Eloísa, D.-G., 2019. Servitization and digitalization in manufacturing: the influence on firm performance. *J. Bus. Ind. Market.* 35, 564–574. <https://doi.org/10.1108/JBIM-12-2018-0400>.
- Mariani, M., Borghi, M., 2019. Industry 4.0: a bibliometric review of its managerial intellectual structure and potential evolution in the service industries. *Technol. Forecast. Soc. Change* 149, 119752. <https://doi.org/10.1016/j.techfore.2019.119752>.
- Marra, A., Antonelli, P., Pozzi, C., 2017. Emerging green-tech specializations and clusters – a network analysis on technological innovation at the metropolitan level. *Renew. Sustain. Energy Rev.* 67, 1037–1046. <https://doi.org/10.1016/j.rser.2016.09.086>.
- Marra, A., Carlei, V., Baldassari, C., 2020. Exploring networks of proximity for partner selection, firms' collaboration and knowledge exchange. The case of clean-tech industry. *Bus. Strat. Environ.* 29, 1034–1044. <https://doi.org/10.1002/bse.2415>.
- Massa, S., Testa, S., 2008. Innovation and SMEs: misaligned perspectives and goals among entrepreneurs, academics, and policy makers. *Technovation* 28, 393–407. <https://doi.org/10.1016/j.technovation.2008.01.002>.
- McKinsey Global Institute, 2016. The Age of Analytics : Competing in a Data-Driven World.
- Mercandetti, F., Larbig, C., Tuozzo, V., Steiner, T., 2017. Innovation by collaboration between startups and SMEs in Switzerland. *Technol. Innov. Manag. Rev.* 7, 23–31. <https://doi.org/10.22215/timreview/1125>.
- Mun, C., Kim, Y., Yoo, D., Yoon, S., Hyun, H., Raghavan, N., Park, H., 2019a. Discovering business diversification opportunities using patent information and open innovation cases. *Technol. Forecast. Soc. Change* 139, 144–154. <https://doi.org/10.1016/j.techfore.2018.11.006>.
- Mun, C., Yoon, S., Park, H., 2019b. Structural decomposition of technological domain using patent co-classification and classification hierarchy. *Scientometrics* 121, 633–652. <https://doi.org/10.1007/s11192-019-03223-8>.
- Nambisan, S., 2017. Digital entrepreneurship: toward a digital technology perspective of entrepreneurship. *Enterpren. Theor. Pract.* 41, 1029–1055. <https://doi.org/10.1111/etap.12254>.
- Nambisan, S., Wright, M., Feldman, M., 2019. The digital transformation of innovation and entrepreneurship: progress, challenges and key themes. *Res. Pol.* 48, 103773. <https://doi.org/10.1016/j.respol.2019.03.018>.
- Onufrey, K., Bergek, A., 2020. Second wind for exploitation: pursuing high degrees of product and process innovativeness in mature industries. *Technovation* 89, 102068. <https://doi.org/10.1016/j.technovation.2019.02.004>.
- Oztemel, E., Gursev, S., 2020. Literature review of Industry 4.0 and related technologies. *J. Intell. Manuf.* 31, 127–182. <https://doi.org/10.1007/s10845-018-1433-8>.
- Park, T.-Y., Lim, H., Ji, I., 2018. Identifying potential users of technology for technology transfer using patent citation analysis: a case analysis of a Korean research institute. *Scientometrics* 116, 1541–1558. <https://doi.org/10.1007/s11192-018-2792-9>.
- Plessis, M.d., Boon, J.A., 2004. Knowledge management in eBusiness and customer relationship management: South African case study findings. *Int. J. Inf. Manag.* 24, 73–86. <https://doi.org/10.1016/j.ijinfomgt.2003.10.002>.
- Reinartz, W., Wiegand, N., Immschloss, M., 2019. The impact of digital transformation on the retailing value chain. *Int. J. Res. Market.* 36, 350–366. <https://doi.org/10.1016/j.ijresmar.2018.12.002>.
- Rindfleisch, A., O'Hern, M., Sachdev, V., 2017. The digital revolution, 3D printing, and innovation as data. *J. Prod. Innovat. Manag.* 34, 681–690. <https://doi.org/10.1111/jpim.12402>.
- Rohrbeck, R., 2010. Harnessing a network of experts for competitive advantage: technology scouting in the ICT industry. *R D Manag.* 40, 169–180. <https://doi.org/10.1111/j.1467-9310.2010.00601.x>.
- Saebi, T., Foss, N.J., Linder, S., 2019. Social entrepreneurship research: past achievements and future promises. *J. Manag.* 45, 70–95. <https://doi.org/10.1177/0149206318793196>.
- Samuelson, P., Lager, T., 2019. Managing product variety under operational constraints: a process-industrial outlook. *J. Bus. Chem.* 2, 134–147. <https://doi.org/10.17879/64159695138>.
- Shepherd, D.A., 2015. Party On! A call for entrepreneurship research that is more interactive, activity based, cognitively hot, compassionate, and prosocial. *J. Bus. Ventur.* 30, 489–507. <https://doi.org/10.1016/j.jbusvent.2015.02.001>.
- Shin, J., Park, Y., 2010. Evolutionary optimization of a technological knowledge network. *Technovation* 30, 612–626. <https://doi.org/10.1016/j.technovation.2010.04.004>.
- Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Manag. Sci.* 51, 756–770. <https://doi.org/10.1287/mnsc.1040.0349>.
- Snihr, Y., Wiklund, J., 2019. Searching for innovation: product, process, and business model innovations and search behavior in established firms. *Long. Range Plan.* 52, 305–325. <https://doi.org/10.1016/j.lrp.2018.05.003>.
- Song, C.H., Elvers, D., Leker, J., 2017. Anticipation of converging technology areas — a refined approach for the identification of attractive fields of innovation. *Technol. Forecast. Soc. Change* 116, 98–115. <https://doi.org/10.1016/j.techfore.2016.11.001>.
- Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. *Res. Pol.* 35, 994–1017. <https://doi.org/10.1016/j.respol.2006.05.002>.
- Stayton, J., Mangematin, V., 2016. Startup time, innovation and organizational emergence: a study of USA-based international technology ventures. *J. Int. Entrep.* 14, 373–409. <https://doi.org/10.1007/s10843-016-0183-y>.
- Stock, T., Seliger, G., 2016. Opportunities of sustainable manufacturing in industry 4.0. *Procedia CIRP* 40, 536–541. <https://doi.org/10.1016/j.procir.2016.01.129>.
- Storm, P., Lager, T., Samuelsson, P., 2013. Managing the manufacturing-R&D interface in the process industries. *R D Manag.* 43, 252–270. <https://doi.org/10.1111/radm.12010>.
- Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F., 2018. Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Technol.* 94, 3563–3576. <https://doi.org/10.1007/s00170-017-0233-1>.
- Trott, P., Simms, C., 2017. An examination of product innovation in low- and medium-technology industries: cases from the UK packaged food sector. *Res. Pol.* 46, 605–623. <https://doi.org/10.1016/j.respol.2017.01.007>.
- Venkatesh, V., Davis, F.D., 2000. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manag. Sci.* 46, 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>.
- Verhoeven, D., Bakker, J., Veugelaers, R., 2016. Measuring technological novelty with patent-based indicators. *Res. Pol.* 45, 707–723. <https://doi.org/10.1016/j.respol.2015.11.010>.
- Wang, B., Liu, H., Song, J., 2016. SaaS-based enterprise application integration approach and case study. *J. Supercomput.* 72, 2833–2847. <https://doi.org/10.1007/s11227-016-1625-y>.
- Williams, C., Lee, S.H., 2009. Resource allocations, knowledge network characteristics and entrepreneurial orientation of multinational corporations. *Res. Pol.* 38, 1376–1387. <https://doi.org/10.1016/j.respol.2009.05.007>.
- Wu, C.-Y., Mathews, J.A., 2012. Knowledge flows in the solar photovoltaic industry: insights from patenting by Taiwan, Korea, and China. *Res. Pol.* 41, 524–540. <https://doi.org/10.1016/j.respol.2011.10.007>.
- Xu, L., Da, Duan, L., 2019. Big data for cyber physical systems in industry 4.0: a survey. *Enterprise Inf. Syst.* 13, 148–169. <https://doi.org/10.1080/17517575.2018.1442934>.
- Yang, K.-C., Shih, P.-H., 2020. Cognitive age in technology acceptance: at what age are people ready to adopt and continuously use fashionable products? *Telematics Inf.* 51, 101400. <https://doi.org/10.1016/j.tele.2020.101400>.
- Yao, L., Ge, Z., 2018. Big data quality prediction in the process industry: a distributed parallel modeling framework. *J. Process Contr.* 68, 1–13. <https://doi.org/10.1016/j.jprocont.2018.04.004>.
- Yeung, J.F.Y., Chan, A.P.C., Chan, D.W.M., 2012. Defining relational contracting from the Wittgenstein family-resemblance philosophy. *Int. J. Proj. Manag.* 30, 225–239. <https://doi.org/10.1016/j.ijproman.2011.06.002>.
- Zaborek, P., Mazur, J., 2019. Enabling value co-creation with consumers as a driver of business performance: a dual perspective of Polish manufacturing and service SMEs. *J. Bus. Res.* 104, 541–551. <https://doi.org/10.1016/j.jbusres.2018.12.067>.
- Zhang, Y., Abbas, H., Sun, Y., 2019. Smart e-commerce integration with recommender systems. *Electron. Mark.* 29, 219–220. <https://doi.org/10.1007/s12525-019-00346-x>.