

Prytkova, Ekaterina

Working Paper

ICT's wide web: a system-level analysis of ICT's industrial diffusion with algorithmic links

Jena Economic Research Papers, No. 2021-005

Provided in Cooperation with:

Friedrich Schiller University Jena, Faculty of Economics and Business Administration

Suggested Citation: Prytkova, Ekaterina (2021) : ICT's wide web: a system-level analysis of ICT's industrial diffusion with algorithmic links, Jena Economic Research Papers, No. 2021-005, Friedrich Schiller University Jena, Jena

This Version is available at:

<https://hdl.handle.net/10419/243076>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

*ICT's Wide Web: a System-Level
Analysis of ICT's Industrial Diffusion
with Algorithmic Links*

Ekaterina Prytkova

JENA ECONOMIC RESEARCH PAPERS · # 2021-005

The JENA ECONOMIC RESEARCH PAPERS
is a publication of the Friedrich Schiller University Jena, Germany (www.jecon.de).

ICT's Wide Web: a System–Level Analysis of ICT's Industrial Diffusion with Algorithmic Links

Ekaterina Prytkova

Friedrich Schiller University Jena, Department of Economics and Business Administration

e.prytkova@uni-jena.de

March 16, 2021

Abstract

This paper seeks to contribute to the understanding of diffusion patterns and relatedness within ICT as a technology system in the EU28 region. Considering ICT as a technology system, first, I break down ICT into a set of distinct technologies employing OECD and WIPO classifications. Then, using text analysis and the Algorithmic Links with Probabilities method, I construct industry–technology links to connect industries with ICT and track ICT's diffusion over the period 1977-2020. The analysis highlights the heterogeneity of the technologies that constitute the ICT cluster. As not all ICTs are pervasive and not all ICTs are key technologies, this leads to differences in industry reliance on them. The results indicate that the ICT cluster shows signs of a “phase transition”, passing the phase of building bulk elements of the infrastructure and around the 2000s entering the phase of working on the functionality for business applications deployment and users' convenience. This transition is marked by the surging relevance of ICT technologies such as mobile communication, information analysis, security, and human interface. Studying the ICT as a cluster allows putting each ICT technology in context to compare them in relative terms; this is especially important for the discussion of novel and fast–growing technologies such as Artificial Intelligence (AI). Concerning the structure of industry reliance on the ICT cluster, ICT's penetration is characterized by increasing scope but unevenly distributed scale; depending on the industry and the distinct ICT technology the intensity of their connections varies significantly. Remarkably, looking closer at AI technologies, in line with the current literature, a wide array of “shallow” connections with industries is revealed. Finally, I calculate relatedness metrics to estimate proximity among ICT technologies. The analysis reveals differences in the underlying knowledge base among the overwhelming majority of the ICT technologies but a similar structure of their application base.

Keywords: ICT; algorithmic links; artificial intelligence; relatedness; industry–technology nexus

JEL Classification: O33, O52, O14

I am grateful to Frederique Bone (SPRU, University of Sussex), Simone Vannuccini (SPRU, University of Sussex), Holger Graf (Friedrich Schiller University Jena), Ivan Svetunkov (Lancaster University) and participants of the workshop on The Economics and Management of AI Technologies (BETA, University of Strasbourg) for useful comments and fruitful discussions.

1 Introduction

The *system* or *cluster* of technologies labeled Information and Communication Technologies (ICT) came to the attention of economists in the late 1980s due to a mismatch between expected and de facto productivity dynamics and have been studied extensively since then. The intensive investments into ICT fueled the development and production of new IT capital and the expansion of ICT application. Eventually, catering to heterogeneous demand needs, ICT evolved into a complex and interconnected system of technologies that forms an infrastructure with a variety of applications based on it. However, in the literature ICT is very often considered as a monolith at a coarse, aggregate level, while in fact it is composite. This makes it difficult to isolate the effects (both benefits and failures) of ICT diffusion and understand how they are achieved. Firms, entrepreneurs and policy-makers cannot utilize information on ICT at such aggregated level for strategy and policy design. This leads to coordination failures between ICT supply and demand and makes the development of ICT myopic or haphazard ([Bresnahan, 2019b](#)).

In this paper, I adopt a systemic approach to ICT by considering industrial diffusion of a set of distinct ICT technologies, each separately as well as in relation to each other within the ICT cluster. More precisely, the analysis estimates the scale and scope of industrial connections for each distinct ICT technology and their dynamics. This reveals directions of development by identifying those ICT technologies that intensify connections with industries by moving closer to the center of the knowledge base and those that experience an exploration phase acquiring new industrial connections. The consideration of the ICT cluster allows putting each ICT technology in context to compare scale and scope of their diffusion not only based on individual growth rates but also in relative terms; this is especially important for the discussion of novel and fast-growing technologies such as Artificial Intelligence (AI). Then, I estimate the relatedness among ICT technologies based on their co-occurrence in the knowledge space and in industries to identify the dimension in which ICT technologies are proximate. The study seeks to identify patterns in the dynamics of industrial penetration by ICT over the period 1977–2020 among 28 EU member states. This helps to estimate the modern state of ICT diffusion and put it in historical perspective.

Methodologically, the ICT cluster is captured combining the new ICT taxonomy by OECD ([Inaba & Squicciarini, 2017](#)) and the PATENSCOPE AI index ([WIPO, 2019](#)). The resulting taxonomy used in the paper aggregates patents' IPC technological classes into 13 distinct ICT technologies including AI. Economic activities are represented by 74 industries at the 2-digit level from International Standard Industrial Classification of All Economic Activities rev.4 (ISIC). The primary connections between industries and ICT technologies are established through extraction of keywords from ISIC industrial descriptions and their subsequent search in patents' titles and abstracts. Further refinement of these connections with the Algorithmic Links with Probabilities method ([Lybbert & Zolas, 2014](#)) produces the final ICT technology–industry matrices, one for each subperiod. As these matrices are essentially bimodal networks, several network metrics are applied to analyze the structure and dynamics of industry reliance on ICT. Finally, relatedness indicators provide insights into the overlap of knowledge and application bases among ICT technologies.

This paper contributes to several literature strands. Given that the analysis is set at the

system level and investigates the linkages within the cluster and beyond, this paper belongs to the strand of research on technology systems and the industrial connections that they create (Freeman, 1994; Perez, 2010). It fills the gap of empirical studies that operationalize the concept of technology system applied to ICT. Unlike studies that consider ICT as a monolith and estimate its impact on a set of industries (one technology to many industries), for example, through the lens of General Purpose Technologies (GPTs) (Basu & Fernald, 2007; Castellacci, 2010), this paper constructs a bimodal technology–industry network connecting a number of well–defined ICT technologies with many industries to uncover the industry portfolio corresponding to each ICT technology and decompose pervasiveness of the whole ICT cluster.

The paper builds on studies that focus on the economic impact of ICT (Brynjolfsson & Hitt, 2000; Van Ark et al., 2003; Brynjolfsson et al., 2019) and traces back the source of this impact to particular ICT technologies. By offering an estimation of pervasiveness, this study contributes to the literature on GPTs and their identification (Bekar et al., 2018). Another contribution of this research lies in the field of sectoral patterns of innovation (Malerba, 2002; Castellacci, 2008) by showing the structure of industrial connections through shared technological knowledge base. In this context, the relatedness indicators used in this paper represent an instantiation of research on the principle of relatedness (Hidalgo et al., 2018). As the analysis takes a closer look at AI technologies, the study is a contribution to the Economics of AI (Agrawal et al., 2019b). In particular, to the best of the author’s knowledge, this work is the first to study together AI diffusion among the industries’ knowledge base and the complementarity of AI with other ICT technologies. Finally, considering the methods used, the paper is also an application of the “text-as-data” approach (Gentzkow et al., 2019) to technological dynamics.

This study can inform both technological and economic perspectives. From a technology–centered perspective, it offers a fine–grained estimation of ICTs’ pervasiveness and diffusion patterns. From an economic perspective, the study pinpoints potential loci of adoption externalities by identifying connected ICT technologies. Establishing influential ICT technologies for each industry provides insights on the technological regimes they induce and consequently on the economic conditions such as innovation opportunities, entry dynamics and market concentration, appropriability, and eventually commercial value and price of products and services. Finally, the industry–technology mapping presented here can help policy–makers in identifying related markets that rely on the same technologies even across industrial boundaries as well as technologies related through the same industries–applications; this is especially relevant for the regulation of merger and acquisition applied to digital markets (see, for example, Federico et al. (2020); Morton & Dinielli (2020)).

The paper proceeds as follows. Section 2 lays out chronologically the changing focus of ICT research, from macro level productivity dynamics to micro level changes in organisational routines and occupations within firms; from a coarse notion of ICT through multiplex networks of economic complexity to case studies of particular technologies. At the end of this Section, the systemic nature of the ICT cluster emerges as a crucial feature characterizing ICT. This frames the rationale behind the construction of the industry–technology mapping described in Section 3, with a further focus on the connections of the ICT cluster: with industries and among the distinct ICT technologies. Section 4 discusses the results of the analysis on the within

ICT cluster structure, the estimation of the pervasiveness of particular ICT technologies, and their relatedness through knowledge and application bases. Special attention is devoted to AI technology with a deeper dive into the AI's connections with other ICT. Section 5 concludes.

2 Many Faces of ICT: from Productivity Paradox of the 80s to Modern Technology System

It has been many decades since ICT diffuse inside the economies transforming and creating new markets, business models, and jobs. ICT are enabling technologies engaging in coinvention with application industries to address market demands and organizational supply processes (Bresnahan & Yin, 2017). ICT have induced an encompassing process of digitalization that led to a restructuring of the socio-economic life and redomaining the economy around digital infrastructures.¹ In retrospect, the impact of ICT is immense and visible but complex and often non-linear in how it propagates. The discussion on ICT came across in the context of the first “productivity paradox” or “productivity puzzle” expressed in 1987 by Roach (1987) and Solow (1987) as surging growth of investment in Information Technologies (IT), in particular in computerization, was coupled with retarding growth of productivity. Put simply, the rationale behind attention to ICT is their expected enhancing effect on resource (capital and labor) productivity which at the time was not observed, creating a mismatch between expected increasing and de facto slowing down productivity growth. Eventually, resource productivity and its rate of growth define feasible production volumes hence market expansion and in the end economic growth. Given that productivity dynamics reverberates at economic growth, the first productivity slowdown that lasted until mid-1995 spurred debates on its origins. To investigate this oddity, scholars went from macro trends down to meso and micro data on industries and firms; these studies suggested several potential explanations for observed phenomenon such as an offsetting effect while IT capital substituted for non-IT capital (Dewan & Min, 1997), firm heterogeneity (Brynjolfsson & Hitt, 1995), and mismeasurement (Diewert & Fox, 1999). Indeed, the period from mid-1990s until mid-2000s has been characterized by productivity acceleration in the US, suggesting delayed but substantial contribution of the ICT-producing sectors to aggregate productivity growth and delayed returns on investment in IT capital by ICT-using sectors due to implementation lags associated with learning, complementary capital accumulation and reorganization (Jorgenson et al., 2003; Basu & Fernald, 2007; Corrado et al., 2007; Bresnahan et al., 2002). Timmer & Van Ark (2005) conduct a comparative study of the EU and the US, regions with not completely synchronized business cycles so that differences related to the time lags are exposed. They find that the contribution to the aggregate productivity by ICT-manufacturing sectors and overall ICT capital deepening are the two factors that explain almost fully the US's lead over the EU in labor productivity growth. While the discussion on the contribution of ICT to economic growth was still ongoing, the second productivity slowdown started around mid-2000s (Fernald, 2015; Syverson, 2017), fueling a new wave of debates between techno-optimists and techno-pessimists. These two research strands differ in predictions

¹For example, in 2018 intangible assets account for 84% of value of S&P500 companies including software code and licenses, data and databases; top-5 largest global companies by market capitalization are digital giants Apple, Alphabet, Microsoft, Amazon, Facebook (Gonzalez & Ponemon, 2019).

for the future of productivity dynamics but they all acknowledge the role ICT played initiating structural changes in production processes, workplace, labor demand, and contributing to overall economic growth (Bresnahan & Yin, 2017).

To capture the profound transformations ICT induced and mechanisms through which these transformations unfold, scholars employ different approaches: (i) inductive inference analyzing the changes at the micro and meso level e.g. demand for skills and income distribution (Autor et al., 1998; Michaels et al., 2014), organizational routines and structures (Brynjolfsson & Hitt, 2000), novel products and services (Bakos & Brynjolfsson, 2001), etc.; (ii) deductive analysis assuming a mechanism that potentially produces a set of observed stylized facts and/or state of the economy e.g. modeling new production factors experimenting with a production function, structural models and simulation and/or more technology-centered framework of General Purpose Technologies (GPT) (Castellacci, 2010; Corrado et al., 2009; Basu et al., 2003; Guerrieri & Padoan, 2007). In both (i) and (ii), ICT is very often considered as a monolith, a coarse notion of all information technologies with differentiation between ICT-using and ICT-producing firms and industries. However, the technology under consideration does make a difference. Given that, another approach to capturing ICT transformations is (iii) studying up close the diffusion of a particular technology considering a larger set of factors including (a) *supply-side* and (b) *technological* ones along with demand-side and economic factors such as preferences and price. As argued by Rosenberg (1972), (a) “the rate at which new technologies replace old ones will depend upon the speed with which it is possible to overcome an array of supply side problems” and (b) “better understanding of the timing of diffusion is possible by probing more deeply at the technological level itself, where it may be possible to identify factors accounting for both the general slowness as well as wide variations in the rate of diffusion.”.

Indeed, at the level of individual technology, the economic value is endogenous to the technological function(s) a particular technology can perform, and how exactly this technology executes the function(s): as infrastructure or network vs fully-fledged component or stand-alone product or service (for example, mobile telecommunication network vs integrated circuit). Moreover, different technologies induce different technological regimes that form around them; in turn, a technological regime defines the environment for innovating agents: opportunity and appropriability conditions, properties and channels of transmission of technological knowledge. These conditions implied by the technological regime reverberate to innovation patterns as well as firm size and entry-exit dynamics at the supply side, and foster industrial dynamics and evolution at various rates and directions (Malerba & Orsenigo, 1997). At the same time, “individual technologies are not introduced in isolation. They enter into a changing context that strongly influences their potential and is already shaped by previous innovations in the system.” (Perez, 2010, p.188). These previous innovations might be instantiations of the same technology illustrating path dependency (David, 2007) within one technological trajectory (Dosi, 1982) as well as of another related technology in the *technology system* (Freeman, 1994) capturing dynamic interrelatedness among technologies. Indeed, Freeman (1994) stresses the systemic aspect of technological diffusion. All this applies to ICT as well, hence the ICT cluster can be considered a technology system. Studying ICT as a system of interrelated technologies in connection to industries can provide a better understanding of their diffusion patterns and impact. For

example, several ICT-producing industries, being connected via complementary technologies, might engage in synergistic interactions (Steinmueller, 2002) through, for example, (indirect) networks with adoption externality (Church & Gandal, 2005). In the field of ICT, a famous case of strategic exploitation of such positive externalities is the so-called Wintel standard formed between Microsoft's OS and Intel's processors (Takahashi & Namiki, 2003).

ICT technology system spans over a wide range of industries and hence constitutes a part of their technological knowledge base. ICT carved a slot in the knowledge space becoming a distinct technology system through the tortuous process of upstreaming and technological convergence (Rosenberg, 1963); a complex infrastructure has been built gradually around the function of handling information (Steinmueller, 1996; Greenstein, 2019) performed within commercial as well as military and scientific applications. An increasing number of application industries incorporates ICT in the knowledge base tying their business models, production processes and overall development to the technical progress and manufacturing of ICT goods. On the one hand, this creates an inflow of investments in ICT-producing industries, boosting their growth. On the other hand, numerous application sectors have heterogeneous preferences about the pace of production and performance of ICT capital. Locked inside the so-called dual inducement mechanism (Bresnahan & Trajtenberg, 1995), for example, the semiconductor industry adopted Moore's law as the main roadmap to sustain the demands of its applications and control the pace of development. Nevertheless, preemption strategies and capacity races (Steinmueller, 1992) in producing the next generation of chips have been always present in the industry, contesting the established roadmap. Besides pressures from the heterogeneity of ICT-using industries, a failure to recognize the economic value of ICT application might occur as well; technology-provided opportunities for commercial applications are not always obvious *ex ante* which creates the coordination problem between technical progress in ICT and the technical progress in its applications (Bresnahan, 2019b).

The multiplicity of linkages between economic activities and technologies is highlighted in the literature on regional development and economic complexity (Balland et al., 2019). Such studies represent the path of regional growth through a bimodal network that connects local capabilities² and diversity of economic activities in the region (Hidalgo & Hausmann, 2009). Establishing these connections serves two purposes: (i) it uncovers the correspondence between products or industries (economic activities) and the required knowledge or inputs (capabilities) and (ii) reveals related economic activities that require similar capabilities. The latter is measured through various relatedness metrics (Hidalgo et al., 2018). For example, two industries or products can be considered related if they rely on similar technological knowledge (Breschi et al., 2003; Balland et al., 2019), labor skills (Neffke et al., 2011, 2018), or input-output structure (Essletzbichler, 2015). In dynamic perspective, it helps to explain the entry probability of a new economic activity in a spatial unit based on existing local capabilities (industrial change), and the consequent evolution of local capabilities (structural change) in response to the entrance of new industries (Neffke et al., 2011). In sum, in the field of economic complexity, this circular impact between industries and technological knowledge base as an instantiation of capabilities illustrates the underlying mechanism of regional diversification and growth (Frenken et al., 2007). The

²Capabilities is an umbrella concept that comprises resources, institutional framework, human capital and knowledge (Maskell & Malmberg, 1999).

current study shares this rationale and introduces complexity through the constructed industry–technology mapping with relatedness metrics applied to it. Unlike the economic complexity literature that conventionally focuses on economic activities (industries, products), this paper focuses on ICT technologies as part of the technological knowledge base in the EU28 region to study the diffusion of the ICT cluster among industries.

In sum, ICT constitutes a technology system or cluster, has a transformative effect on economic activities that incorporated ICT in their technological knowledge base, and enables industrial synergies. Referring to ICT as a *cluster* or a *system* is crucial because it stresses the composite nature of ICT and exposes the rationale to consider it as a set of interrelated yet distinct technologies. In this paper, I adopt this systemic approach to ICT and claim that within the ICT cluster technologies are heterogeneous in their nature, leading to uneven scale and scope of adoption among industries. In other words, not all ICTs are pervasive, not all ICTs are key technologies³. A more fine–grained consideration of the ICT cluster can reveal the structure of its pervasiveness by identifying ICT technologies that experience (a) increasing scale of penetration by deepening the connection with industries or (b) increasing scope of application by creation of new applications/markets or both. The analysis of relatedness is aimed at uncovering the dimensions along which heterogeneous ICT technologies are proximate. Altogether, this draws a more complex and up–to–date picture of the ICT cluster.

3 Methodology: Constructing Industry–Technology Mapping

To conduct the analysis, I employ OECD and WIPO patent–based classifications to break down ICT into a set of technologies, and using text mining and probabilistic matching, construct technology–industry nexus tracking its development over time. This nexus is a dynamic mapping between economic activities represented by industries and the set of ICT technologies. Estimating how distinct ICT technologies penetrate industries’ knowledge base in dynamics and looking deeper into the origin of these relations provides a more accurate and meaningful view on ICT diffusion. The approach I have employed to construct industry–technology mapping is based on the Algorithmic Links with Probabilities (ALP) method proposed by [Lybbert & Zolas \(2014\)](#). This method allows establishing industry–technology connections in two essential steps: (i) connect industries and patents via the search of keywords extracted from industries’ descriptions in patent’s abstract and title, and then, based on these links through patents, (ii) connect industries with patents’ technological classes that belong to the ICT taxonomy. Further refinement of obtained industry–technology frequency matches implies a transformation of simple cross–tabulation values into Bayesian probabilities.

The ALP method has a number of advantages in comparison with other industry–technology concordances (e.g. Yale Technology Concordance (YTC) ([Kortum & Putnam, 1997](#)) and DG Concordance ([Schmoch et al., 2003](#)) as it is: (i) modifiable — new keywords, industries, or technologies can be added and linkages easily recalculated without reconstructing the whole mapping from scratch; (ii) dynamic — over time the industry–technology linkages can emerge or disappear; the method is dynamic as it allows construction of the mapping for any defined period; (iii)

³“Key technologies are defined as holding a central position within the knowledge base.”([Graf, 2012](#))

scalable — technology and industry classifications employed in the study each has several levels e.g. 4-, 3-, 2-digit level; the ALP method can be applied to any combination of classifications’ levels; once industry–technology connections are calculated at chosen levels, further aggregation along each classification is possible: for example, 2–digit level (ISIC divisions) can be further aggregated up to 1–digit (ISIC sections) level by simple summation of calculated connections for nested industries. Overall, the ALP method employed for the purposes of this paper helps to connect meaningfully industries to large but distinct ICT classes, not to a particular technology confined in a patent.

Industrial classification. The first step of the ALP method requires the extraction of keywords that characterize the economic activity the industry carries out. The description of industries comes from the United Nations’ International Standard Industrial Classification of All Economic Activities Revision 4 (ISIC Rev.4) (UN, 2008). It is a suitable choice of meso–level classification that spans over all sectors of the economy from the primary sector with agriculture and raw materials production to the tertiary sector of services such as consultancy, advertising, research, etc. The choice of ISIC taxonomy’s depth fell on 2–digit level because this level at the same time allows for a sufficient amount of text description per industry and produces a fine–grained matching between industries and technologies. Overall, there are 74 industries included in the analysis.

Keywords extraction. The purpose of keyword extraction is to create a set of characteristic tokens to represent each industry. The text corpus used for keyword extraction is ISIC rev.4 industrial description with industries at the 2–digit level (division) including their nested levels (3–digit group and 4–digit class) treated as separate documents. The choice to go for the characteristic tokens or phrases, so-called n–grams, is motivated by the need to balance between type I and type II errors: exclusion of useful tokens (false positive) and inclusion of distorting tokens (false negative). The latter means that some words can have multiple and sometimes quite distant meanings while indeed being actively used in a particular economic activity. Thus, the exclusion of such words from the keywords set would harm the representation of an industry while inclusion would confuse/conflate several industries. Bigrams help to solve this conundrum because they can consist of separately ambiguous words and by combining them create a phrase with a more specific meaning that allows attributing it to a particular industry unequivocally and not losing an important word. Table 1 illustrates the described principle with some examples.

| Word | Bigram | ISIC |
|-----------|-------------------------|------|
| equipment | irrigation equipment | 16 |
| | communication equipment | 26 |
| | signaling equipment | 27 |
| | freezing equipment | 28 |
| | dental equipment | 32 |
| | optical equipment | 33 |
| plant | forage plant | 11 |
| | plant propagation | 13 |
| | power plant | 42 |
| | nuclear plant | 71 |
| | sewage plant | 81 |

Table 1: Disambiguation with bigrams

Breaking down the text into tokens leads to a large list of single words and bigrams with many of them being redundant. To illustrate the process with an example, one sentence of n words turns into a list with n single words and $(n - 1)$ bigrams. Removal of n -grams that contain stop-words, such as articles, forms of the verb *to be*, etc. accounts for only a fraction of cleansing of this list. To select meaningful n -grams after the removal of stop-words, I have applied two techniques (i) Part-of-Speech tagging (PoS) and (ii) calculation of term frequency-inverse document frequency statistic (TF-IDF).

The TF-IDF statistic is a composite indicator that helps to construct a broad representation of what a document is about. Precisely, it estimates how important is an n -gram to a document based on its occurrence frequency within and between documents. The first component, term frequency (TF), is a simple frequency of an n -gram within a document that shows how often the n -gram occurs in the document. The second component, inverse document frequency (IDF), divides the total number of documents by the number of documents that contain the n -gram which reaches its minimum (equals to 1) when the n -gram is found in all documents and its maximum (equals to $\log(n)$) when the n -gram belongs to only one document. The product of these two parts form the TF-IDF statistic that is high for n -grams that are frequent within one document but is not common for the rest documents. Calculation of TF-IDF is possible for n -gram of any length. In application to ISIC description of industries, the notion of document is equivalent to 2-digit level industry description with all nested 3- and 4-digit level descriptions. More generic words, like already used example of word *equipment*, go down in the ranking because they can occur in many 2-digit industries at the same time.

The PoS tagging is a Natural Language Processing (NLP) technique that helps to identify word's part of speech (noun, verb, adjective, gerund, etc.) given the context of the text. Consideration of the context in identifying word's part of speech is important because of the coincidence between forms of different parts of speech (e.g. *to fish* and *a fish*) which creates ambiguity which tag to attach to a word. Therefore, the PoS method is applied to the raw text to identify contextually the part of speech for every word in the text. As a result, the text is transformed into a lexicon where every word has a corresponding PoS tag. Only then the text is broken down into n -grams which should be sorted in the following way. First, nouns both singular and plural and gerund parts of speech are selected from the words list as potential candidates for keywords. Second, with bigrams the criterion of selection is put on each word separately to create a meaningful combination: the first word can be a noun, gerund or adjective while the second can be still only a gerund or noun.

Stop-words removal and application of these two techniques to the ISIC industrial descriptions provide each type of n -gram (word, bigram) with its PoS tag and TF-IDF statistic. This treatment cleanses the initial list of n -grams to almost purely characteristic phrases. However, the number of key n -grams per industry is not evenly distributed across industries because the original description can consist of few short lines that result in a few key n -grams to extract. To tackle this problem and reduce type I error for industries represented with a small number of tokens, the set of tokens was expanded using two methods: (i) synonyms search and (ii) vocabulary expansion. The first method, expansion through synonyms, uses PATENTSCOPE's

Cross Lingual Expansion.⁴ This tool provides synonyms found in patents' texts based on selected technological domains. The same tool can help with finding synonyms for words with many meanings that differ conditional on the context hence it also helps to act upon type II error, the same problem as the usage of bigrams does. To illustrate the mechanism at work, consider a word *skin* whose lexical connotation changes from textile industry to pharmaceutical one. Thus, for textile, the synonyms can be *leather*, *pelt*, and *hide*, while for pharmaceuticals *derma* can be used.

The second method, vocabulary expansion, uses patents selected during the first round of search of the preliminary list of keywords and extracts additional keywords relevant to the economic activity from titles and abstracts of these selected patents. Consider an industry with k key tokens (both words and bigrams). These k tokens are found in n patents' abstracts and titles. Most of these n patents contain only one token, less contain two tokens and so on; in other words, such histogram is skewed resembling some asymmetric distribution like Pareto or exponential. I select a subset of m patents out of these n where 2 or more tokens are found.⁵ The titles and abstracts of the subset m patents are broken down into tokens as well; stop words removal and TF-IDF are applied to extract additional characteristic tokens. Given that a patent mostly contains information about technology and only a small share of patent's abstract might describe the application related to the economic activity, tokens extracted in this way were manually revised and selected. The final list of key n-grams contains slightly more than 4800 regularized n-grams for 74 2-digit industries.⁶

Algorithmic Links with Probabilities. The link between an industry and a technology is established through searching a key token extracted from the industry's description in the patent's title and abstract. Thus, technology areas that a patent belongs to according to International Patent Classification (IPC) serve as an approximation for technologies that industry might rely on. The sample of patents to construct the mapping is limited to original European patents where either inventor or an applicant is located in one of 28 European countries (EU28) based on the data from the OECD REGPAT 2020 database.⁷ The data on abstracts and titles of EU28 patents is retrieved from the EPO Worldwide Patent Statistical Database (PATSTAT) Spring 2020.

The ALP method is extensively described in Lybbert & Zolas (2014), hence here I will outline only its essence and some important details for this study. Once a key token for the industry is found in the title or abstract of a patent, it also creates a connection between the industry and the IPC classes of the patent.⁸ The IPC classes represent technologies. Therefore, the outcome of the first step of the ALP method is a matrix with a simple count of matching between industries and IPC classes. To transform the raw count into the industry-to-technology ALP concordances, Bayes rule is applied. The resulting ALP concordances are Bayesian probabilities adjusted to account for some technological fields that can be naturally very prolific in patenting

⁴<https://patentscope.wipo.int/search/en/clir/clir.jsf>.

⁵i.e. the full set of patents found with k key tokens is $f(1 \leq x \leq k) = n$ while a subset of patents that contain at least two key tokens out of k is $f(2 \leq x \leq k) = m$ where $m \leq n$

⁶The quantiles for the number of key tokens per industry: $Q_{25} = 34, Q_{50} = 50, Q_{75} = 74$

⁷Consideration of EU28 countries that include the UK is motivated by the UK's participation in the EC and later in the EU since 1973. Given that the time period in the study covers years from 1977 till 2020, it justifies the usage of the data on EU28.

⁸In this study the 4-digit level of IPC classes is taken for the mapping construction.

hence by size compress shares of other, less prone to patenting, technological fields in industry's recipe. Thus, the share W_{ij}^H of technology j (IPC_j) in industry's i ($ISIC_i$) technological recipe is calculated according to the following formula:

$$W_{ij}^H = \frac{Pr(ISIC_i|IPC_j) \times W_{ij}^R/J}{Pr(ISIC_i|IPC_1) \times W_{i1}^R/J + \dots + Pr(ISIC_i|IPC_J) \times W_{iJ}^R/J} \quad j = \overline{1, J} \quad (1)$$

$$W_{ij}^R = \frac{Pr(ISIC_i|IPC_j) \times Pr(IPC_j)}{Pr(ISIC_i|IPC_1) \times Pr(IPC_1) + \dots + Pr(ISIC_i|IPC_J) \times Pr(IPC_J)} \quad (2)$$

where J is a number of IPC classes at the 4–digit level. Inevitably, some rubbish connections can show up due to a large number of abstracts and titles and various key n–grams. To eliminate such connections the cut–off threshold is set at 2% meaning shares W_{ij}^H lower than 0.02 are set to zero and the remaining shares are renormalized to sum up to 1. In sum, the ALP concordance matrix W^H for each period consists of 74 2–digit ISIC industries and 638 technologies at 4–digit level of IPC classes (IPC4). The choice of IPC4 level is motivated by construction of the ICT taxonomy so that it allows application of the ALP method; the next paragraphs discuss it in details.

ICT cluster. The constructed mapping is used to track the diffusion of the ICT cluster over time. The representation of the ICT cluster expressed in patents' IPC classes is constructed for this study combining two taxonomies: (i) the new ICT taxonomy of OECD ([Inaba & Squicciarini, 2017](#)) and (ii) PATENTSCOPE AI Index ([WIPO, 2019](#)). The new ICT taxonomy by OECD provides concordance between 13 ICT classes such as High speed network and Mobile communication and various–level IPC classes. This taxonomy is taken as the main structure of the ICT cluster. The modification of the OECD taxonomy concerns its class Cognition and meaning understanding which is turned into a class of Artificial Intelligence (AI). The rationale behind is that original class Cognition and meaning understanding represents a subset of AI while the latter gains place inside the ICT cluster as a distinct class of technologies by entering the active commercial phase and experiencing intensive development and experimentation. First, I have identified that IPC classes which represent AI techniques and AI functions from WIPO's AI Index are found only in three ICT groups of OECD taxonomy: (i) a substantial overlap with class Cognition and meaning understanding, (ii) Imaging and sound technology and (iii) Others. Then, to create AI class on the basis of Cognition and meaning understanding, all WIPO's AI–related IPC classes were excluded from (ii) and (iii) and transferred to (i) so the 13 ICT classes remain mutually exclusive. [Table 2](#) contains the resulting ICT taxonomy.

Overall, 55 4–digit IPC classes constitute the ICT cluster. Limiting the constructed mapping to these 55 IPC classes and their subsequent aggregation into 13 ICT classes focuses the attention on the ICT cluster. However, according to the ICT taxonomy, only a fraction of 4–digit IPC class might be related to the ICT cluster. Moreover, the distinction between one or the other ICT group can occur at a deeper level of IPC class than 4–digit; an ICT group can consist of various–level IPC classes at the same time such as 4–digit along with fine–grained 8–digit. In other words, the ICT taxonomy is constructed on varying IPC level.

| ICT class | Description |
|-----------|---------------------------------------|
| cl1 | High speed network |
| cl2 | Mobile communication |
| cl3 | Security |
| cl4 | Sensor and device network |
| cl5 | High speed computing |
| cl6 | Large-capacity and high speed storage |
| cl7 | Large-capacity information analysis |
| cl8 | Artificial Intelligence* |
| cl9 | Human-interface |
| cl10 | Imaging and sound technology |
| cl11 | Information communication device |
| cl12 | Electronic measurement |
| cl13 | Others |

*Created on the basis of class Cognition and meaning understanding by merging with PATENTSCOPE AI Index taxonomy

Table 2: The modified new ICT taxonomy by OECD

For example, in the upper panel of Table 3 distribution of IPC4 class G06K among 13 ICT groups is shown. The division of IPC G06K between ICT classes occurs on a 5- and 6-digit levels and one subclass, G06K21, doesn't belong to ICT at all. Therefore, counting every patent with G06K IPC class as 1 for each ICT group connected with G06K — ICT classes 3, 6, 8, 9, 13 — would inflate the size of each ICT class as not all patents with G06K IPC4 class belong to ICT class 6 or 13. In general, the ALP method uses a single level of IPC to construct concordance, e.g. 4-digit, and cannot use different levels at the same time. For example, a concordance matrix constructed at 5-digit level can be aggregated to 4-digit level, but it cannot contain 4- and 5-digit levels simultaneously. Thus, to keep the ALP concordances on the chosen 4-digit level with respect to the technologies and avoid double counting, the following shares are calculated based on the *sample*. First, the between share, the de facto share of an IPC4 class that belongs to the whole ICT cluster, without division into 13 groups. According to the lower panel of Table 3, in the third period, 99.9% of patents with IPC4 class G06K belong to the ICT cluster i.e. patents' IPC classes (4-digit and longer) match with IPC classes listed in the ICT taxonomy. This means that in the EU28 sample, there is only 0.1% of patents that belong to non-ICT subclass G06K21 in the third period. Second, the within share, the de facto share of an IPC4 class that relates to one of the 13 ICT classes. In other words, the between share is now further decomposed into 13 shares, each for one ICT class, that are renormalized to sum up to 1. In the example of G06K class, the 99.9% between share is decomposed into within ICT cluster shares shown in the lower panel of Table 3. For instance, the 29.8% within share of class 3 Security means that inside the 99.9% between share of IPC4 class G06K there are 29.8% of patents that belong to longer IPC subclass G06K19.

In sum, the IPC4 level for the study is a convenient choice as 4-digit level is the highest among IPC classes used in the ICT taxonomy because combined with the between and within ICT shares, it allows (i) working with a single level of IPC classes to apply the ALP method and (ii) avoid double the counting by tracking precisely the share of the whole ICT cluster in the technological composition of industries and the shares of its 13 classes separately.

Figure 1 shows the stepwise procedure that leads to obtaining industry-ICT class matrices for each period. Starting from extraction of preliminary keywords, the subsequent synonym

| Decomposition of G06K 4-digit class among ICT groups based on the OECD taxonomy | | | | | | | | | | | | | | |
|---|-----|-----|--------|-----|-----|--------|-----|-------|--------|------|------|------|--------|-------------|
| G06K | cl1 | cl2 | cl3 | cl4 | cl5 | cl6 | cl7 | cl8 | cl9 | cl10 | cl11 | cl12 | cl13 | non-ICT |
| | | | G06K19 | | | G06K1 | | G06K9 | G06K11 | | | | G06K15 | G06K21 |
| | | | | | | G06K3 | | | | | | | G06K17 | |
| | | | | | | G06K5 | | | | | | | | |
| | | | | | | G06K7 | | | | | | | | |
| | | | | | | G06K13 | | | | | | | | |
| Within and between shares based on actual patent count in the sample | | | | | | | | | | | | | | |
| G06K | cl1 | cl2 | cl3 | cl4 | cl5 | cl6 | cl7 | cl8 | cl9 | cl10 | cl11 | cl12 | cl13 | Between ICT |
| Within ICT | 0.0 | 0.0 | 29.8 | 0.0 | 0.0 | 19.1 | 0.0 | 0.0 | 48.0 | 0.0 | 0.0 | 0.0 | 3.1 | 99.9 |
| (% of between) | | | | | | | | | | | | | | |

Table 3: The between and within shares of G06K class in the ICT cluster in the third period 2005–2020

and vocabulary expansion result in the final list of 4.8 thousand tokens that characterise 74 industries. Frequency matches between industries and IPC4 patent classes that represent all technologies are further transformed into Bayesian probabilities. The modified ICT taxonomy created based on merged OECD and WIPO classifications is applied to subset ICT technologies (55 IPC4 classes out of 638) obtaining industry–ICT matrices W^H for each period. Finally, the correction with between and within shares is applied to avoid inflation of ICT groups’ sizes.

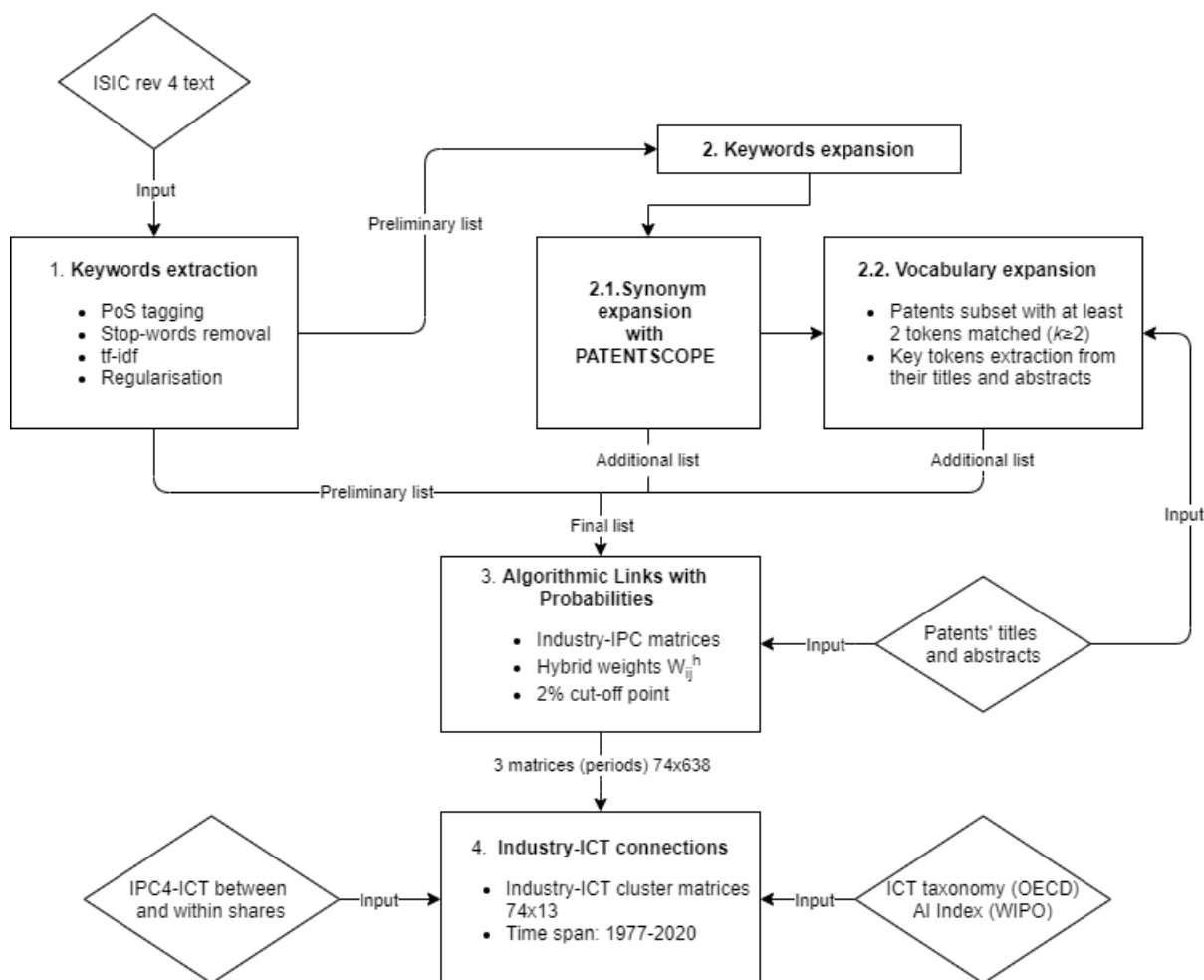


Figure 1: Procedure of obtaining industry–ICT concordance with used techniques and data

Network analysis and relatedness. As a next step, the industry–ICT class matrices are transformed into bipartite networks. In this study, ICT technologies represent a subset of tech-

nological knowledge base i.e. proxy for a part of local capabilities. The set of industries reflects diversity of economic activities in the EU28 region related to ICT technologies. This part of analysis is focused on ICT technologies unlike studies on economic complexity that conventionally focus on economic activities. However, the paper shares the methodological approach and the stance on capability–activity dynamic interrelatedness as an important mechanism of growth.

Network analysis provides a variety of tools to uncover structural properties of a network. For example, flow betweenness centrality (FBC) is used in this paper to rank ICT classes according to their mediation role in connecting industries. It is worth noting that industry–ICT class connections are shares (W_{ij}^H from Equation 1) or probabilities hence the higher the share of a particular ICT class in a technological recipe of an industry, the stronger the connection. This explains the rationale behind the choice of FBC, as it accounts for the strength of connections between an ICT class and all incident industries, and thus shows *a weighted scale of industry reliance on each ICT class*. Nevertheless, in the context of industries–ICT relationship, a node’s degree carries useful information about *the scope of industry reliance on each ICT class*. These two indicators provide further details for a more fine–grained consideration of otherwise homogeneous ICT cluster.

Another dimension to study ICT technologies is their relatedness. As pointed out earlier in the paper, very often ICT technologies are pooled together under an umbrella label of generic ICT or digital technologies. However, not only ICT cluster comprises a wide array of distinct technologies, their alleged similarity can be tested. To do so, I calculate two metrics, technological and application relatedness, to measure the proximity of ICT technologies with regard to the underlying knowledge and application bases respectively.

The *technological relatedness* applied to ICT technologies shows which of them are rooted in similar knowledge base i.e. related through shared knowledge areas. This is measured as a standardized frequency of co–occurrence of ICT classes in patents. If, according to the modified OECD taxonomy, IPC classes of a patent belong to two different ICT classes this represents an instance of co–occurrence; the total number of co–occurrences of two ICT classes is an absolute frequency. However, the absolute frequency of co–occurrence can be a misleading indicator of technological relatedness for two reasons. First, a larger ICT class has a higher potential to co–occur with other ICT classes even by randomness. Second, in absolute terms frequency of co–occurrence of two small ICT classes, for example, i and j , might look negligible in comparison with co–occurrence of two large ICT classes, p and k , and yet represent significant technological relatedness of i and j . Therefore, a test for the randomness that accounts for the size of the ICT classes (number of patents assigned to each of them) must be conducted. Following [Breschi et al. \(2003\)](#), I assume that the frequency of co–occurrence of any two ICT classes, O_{ij} , is a hypergeometric random variable X_{ij} . Thus, ICT classes i and j co–occur in exactly x patents with the following probability, mean and variance:

$$P[X_{ij} = x] = \frac{\binom{R_i}{x} \binom{T - R_i}{R_j - x}}{\binom{T}{R_j}} \quad \mu_{ij} = \frac{R_i R_j}{T}; \quad \sigma_{ij}^2 = \mu_{ij} \left(1 - \frac{R_i}{T}\right) \left(\frac{T}{T - 1}\right) \quad (3)$$

where T is the number of patents in a particular period, R_i and R_j patents belong to ICT class i and j respectively. The standardized frequency of co-occurrence (with zero mean and unit variance) is equal to:

$$\tau_{ij} = \frac{O_{ij} - \mu_{ij}}{\sigma_{ij}} \quad (4)$$

The τ_{ij} statistic takes into account the size of the ICT classes and compares the actual co-occurrence with the expected one as if patents were assigned to ICT classes randomly. If the statistic is significant the null hypothesis of a random relationship between two ICT classes is rejected; the positive statistic $\tau_{ij} > 0$ suggests the presence of technological relatedness of two ICT classes while the negative one $\tau_{ij} < 0$ means two ICT classes occur together even less often than by random assignment.

The *application relatedness* is a correlation that measures how (dis)similar the structure of industry reliance for each pair of ICT classes. There are two significant differences of this novel indicator in comparison with other relatedness metrics used in the literature on economic complexity. First, as the focus is placed on technologies and not on industries, the application relatedness indicates proximity of two technologies through linkages to shared industries i.e. shared application base for the two technologies. Thus, the logic of the application relatedness is reversed with regard to co-occurrence based industry relatedness. For example, [Neffke et al. \(2011\)](#) derive relatedness of industries “from the co-occurrence of products that belong to different industries in the portfolios of manufacturing plants”. Technically, [Neffke et al. \(2011\)](#) calculate product relatedness, and the industry relatedness derived from it is called *revealed relatedness* because it is measured through the intermediate layer of products. However, the overarching logic of industrial relatedness as co-occurrence of industries in plants stands. Reversing this logic, I define application relatedness as co-occurrence of ICT technologies in industries using industry-technology matrices W^H .

To calculate co-occurrence based relatedness for a set of technologies, for example, a technology-to-technology adjacency matrix is used. Each cell of such matrix contains the co-occurrence frequency of two respective technologies in all relevant industries pooled together and each counted as one. This might lead to an incorrect estimation of relatedness between technologies. For example, in a conventional relatedness metric, if two technologies are jointly present in a particular industry, this would add one to the overall co-occurrence count. Such binary count and summation to obtain frequencies do not include useful information respectively on (i) the strength of industry-technology connections and on (ii) the set of particular industries relevant to each of the two technologies. Together (i) and (ii) constitute *a distribution or a structure of industrial connections of each technology*. Thus, the second difference of the suggested application relatedness is that it captures the similarity of distribution of industrial connections for a pair of technologies. By accounting for both (i) and (ii), the metric can capture relatedness of technologies that both have weak connections with industries but the distribution of their connections is similar for the two technologies.

The application relatedness calculates the correlation between two columns that represent

two ICT technologies of the industry–technology matrix W^H .⁹ In each period there are m industries (indexed $i = \overline{1, m}$) that rely on at least one of the 13 ICT classes (indexed $j = \overline{1, n}$ where $n = 13$) (see Appendix A) for some share or weight W_{ij}^H . For example, for ICT class $j = 1$ there are $k < m$ industrial connections¹⁰ $W_{i=\overline{1, k}, j=1}^H = (W_{1,j=1}^H, W_{2,j=1}^H, \dots, W_{k,j=1}^H)$. This vector represents the structure of industry reliance on ICT class $j = 1$. The same vector exists for ICT class $j = 2$ (and any other class in the ICT cluster). Therefore, the application relatedness between class 1 and 2 is calculated as:

$$r_{1,2} = \text{corr}(W_{i,j=1}^H, W_{i,j=2}^H) \quad \forall \quad i = \overline{1, m} \quad (5)$$

Together the two indicators, application and technological relatedness, create four combinations characterizing similarity among ICT technologies. There are shown in the quadrants of Table 4. This scheme provides a useful framework to identify potential loci of adoption externalities and drivers of development for each pair of ICT technologies. As implied by the notions of dynamic interrelatedness and dual inducement mechanism, the development of a particular ICT technology is linked to and influenced by both its related technologies and industries. For example, Bresnahan (2019b) distinguishes between two types of ICT, namely scientific and engineering ICT and commercial and enterprise ICT, and argues that innovation processes occurring in each domain are subject to different factors. In a nutshell, the invention of a scientific and engineering ICT application follows “purely technical requirement” while “[t]he invention of the applications of ICT in much of commercial and enterprise uses necessarily takes the analysis outside “purely technological level”” and follows “visibility” or obviousness of application. Thus, for example, in the first and the forth quadrant, where the technologies are proximate in terms of underlying knowledge, it is easier to find a common technical ground for two ICT technologies and innovate based on available technical possibilities. While in the second quadrant, where ICT technologies exhibit similar strength of connections across shared set of industries but not proximate in knowledge space, the invention might indeed follow “visibility” of commercial value. In general, if a pair of ICT technologies exhibits high application relatedness, one might observe bundling of products and services embodying these two technologies inside a shared set of industries as it is motivated by commercial value. If technologies are not only co–present in an industry but also complementary, that will produce adoption externalities which imply even bigger commercial value but also larger risks (for concrete examples see Simcoe & Watson (2019)). An even further degree of integration between two application–related technologies is mergers and acquisitions among firms producing products and services that have commercial value in bundling and/or exhibit adoption externalities. Therefore, the estimation of application relatedness can inform regulators and policy–makers by identifying related markets even across industrial boundaries connected by technologies that themselves are not related in the knowledge space.

In sum, the rationale behind the construction and usage of the mapping with the ALP method is the dynamic nature of constructed concordances as that allows capturing the changing reliance

⁹Undoubtedly, this is a coarse, linear approximation of the relationship among ICT classes through industries. One can test for non–linear relations using polynomial models or various link functions under the Generalized Linear Model (GLM) framework.

¹⁰The length of all vectors is m as the remaining $m - k$ connections that do not exist between an ICT class and an industry are set to 0.

| | | | | | |
|--------------------------------|-------------|---|---|----------------|-------------------------|
| | | Technological relatedness | | | |
| | | low | high | | |
| Application relatedness | high | Different knowledge, similar industries | Similar knowledge, similar industries | similar | Application base |
| | low | Different knowledge, different industries | Similar knowledge, different industries | | |
| | | different | similar | | |
| | | Knowledge base | | | |

Table 4: Relatedness space with four distinct quadrants

of industries on a set of ICT technologies; some industry–technology connection might decline and some might emerge. On the one hand, the invention and patenting process can last for years and it also takes time for a new technology to enter the knowledge base of an industry, hence it makes sense to use longer time periods to construct ALP concordances. On the other hand, considering too long periods would extinguish the dynamic nature of the concordances. Therefore, I split the whole time period into three subperiods approaching 15 years length which also creates a nearly even distribution of patent sample over these three periods: (i) 14 years: 1977–1990; (ii) 14 years: 1991–2004; (iii) 16 years¹¹: 2005–2020. Finally, the chosen industry and technology levels for which ALP matrices are constructed can be changed by aggregation of matrices’ values instead of their recalculation. For example, the industrial level can be aggregated going from 2– to 1–digit ISIC codes without changing the patents’ IPC level or layering up another industry–IPC concordance for the novel combination of levels.

4 Results and Discussion: Inside the ICT Technology System

The application of the methods outlined in Section 3 allows considering the ICT cluster not as a monolith and look behind the common notion of “pervasive ICT”. First, I will provide initial findings based on a bird’s eye view of the cluster. Then I proceed to a more fine–grained level and analyze the ICT cluster as a collection of distinguished technological classes.

Industrial diffusion of the ICT cluster. In this study, the whole economy of EU28 is represented by 74 industries. The construction of ALP concordances reveals that over the time span from 1977 to 2020 the number of industries relying on the ICT cluster increases from 26 to 36 (see Appendix A), covering almost half of the industries. This means that one third of all economic activities in the first period and nearly half in the last systematically incorporates technologies from the ICT cluster into its knowledge base.¹² Thus, the scope of ICT application grows over time. Nevertheless, the intensity (scale) of reliance on the ICT cluster is distributed unevenly across industries as it is shown in the left panel of Figure 2.

The left panel of Figure 2 plots the share of the whole ICT cluster in the technological

¹¹The last period from 2005 till 2020 includes more years because the patenting activity of the last couple of

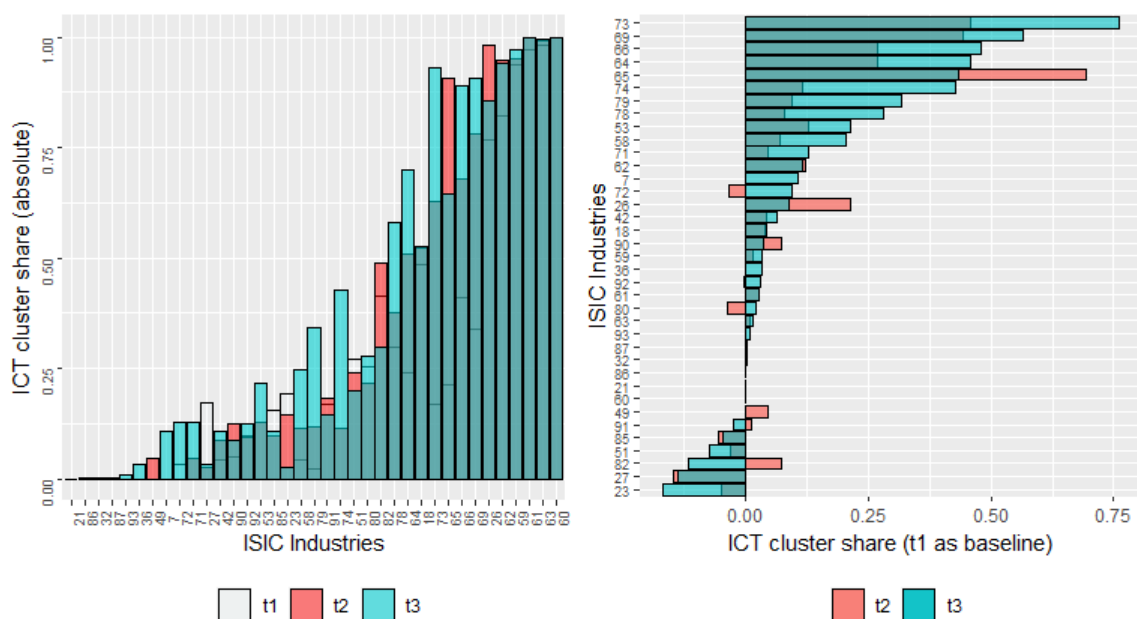


Figure 2: The share of the ICT cluster in technological recipes of industries

Color–period correspondence: period 1 — white, period 2 — red, period 3 — blue. The bars are overlaid with transparency parameter allowing to see all three periods at once. In the left panel, for example, for industry 65 the red bar being the tallest means that in period 2 this industry had the highest share of ICT cluster in its technological composition in comparison with two other periods. In the right panel, period 1 is taken as a baseline by subtracting the share of the first period from the second and third; all bars to the right side of the vertical axis indicate increasing share of the ICT cluster in industries’ recipes while the left ones — decreasing.

recipe of industries. The industries that exhibit the strongest connection with the ICT cluster are not surprising: programming and broadcasting (code 60), information services (63) and telecommunications (61), motion picture, video and television programme production, sound recording and music publishing activities (59), computer programming and consultancy (62), manufacture of computer, electronic and optical products (26). The right panel of Figure 2 displays the first differences of the ICT cluster share in the total technological composition of an industry with the first period as a baseline; it shows that for the majority of industries the connection with the ICT cluster is strengthening. The top-3 industries that experience the highest growth are 73–Advertising and market research (in accord with Anderson (2012)), and 69–Legal and accounting activities followed by the whole Section K: Financial and insurance activities (divisions 64–66). Overall, the majority of industries that exhibit at least some reliance on the ICT cluster belong to services starting from Section H: Transportation and storage onward.¹³

Inner structure of the ICT cluster. The modified OECD taxonomy (see Table 2) allows looking deeper into the ICT cluster distinguishing 13 classes of ICT technologies. Figure 3 helps to conduct an inspection of the ICT cluster composition based on the sample of EU28 patents.

Notably, the leaders of the ranking in the left panel of Figure 3 by absolute share of an ICT class in the ICT cluster — classes High speed network, Information communication device and Large-capacity and high speed storage, — experience a decrease of their shares in the ICT cluster

years is still ongoing and technically they are not fully represented yet

¹²It is worth noting that this estimation is a lower bound because it is based solely on patent data.

¹³The share of industries related to the ICT cluster in Sections H, J, K, M, N, P–R on ICT ranges between 60–100%.

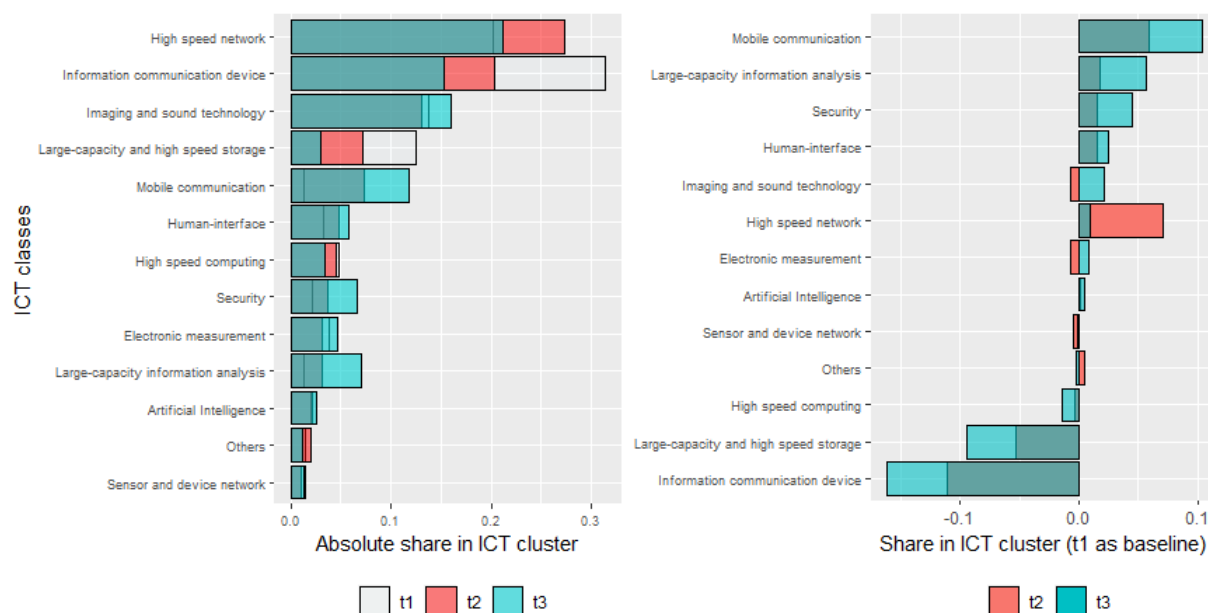


Figure 3: Shares of ICT classes in the cluster

Color-period correspondence: period 1 — white, period 2 — red, period 3 — blue. The bars are overlaid with transparency parameter allowing to see all three periods at once. The left panel plots shares in all three periods while the right panel takes the first period as a baseline subtracting it from the second and third periods. For example, class Mobile communication exhibits its highest share in the ICT cluster in the third period as the blue bar is the tallest. Transparency allows seeing the height of the red bar that indicates share of Mobile communication class in the ICT cluster in the second period.

though keeping strong positions. Instead, a subset of technologies such as Mobile communication, Large-capacity information analysis, Security, and Human interface steadily increase their presence in the ICT cluster indicated by their top positions in the right panel of Figure 3 (first differences with the first period as a baseline). This can be viewed as a sign of structural change experienced by the ICT cluster: from building bulk elements of the infrastructure to transmit data, for example, IDSN and fiber optic networks (Greenstein & Spiller, 1996), Next Generation Networks (Fitchard, 2003), to working on the functionality for numerous applications to make the infrastructure more agile, scalable, secure and affable. The latter can be exemplified with broadband cellular networks such as 4G and 5G, cloud computing, cybersecurity such as HTTPS protocol, virtual assistants, and proliferation of frontend GUIs and VUIs (Graphical and Voice User Interface). In other words, soon after the turn of millennia the completion of the physical infrastructure and its operation processes were mostly over, creating a coherent platform for applications' deployment; the new vector of ICT development is oriented at the improvement of specific aspects of the constructed platform for both businesses and end users. Probably witnessing the end of the first phase of ICT, Nicholas Carr wrote: "While no one can say precisely when the buildout of an infrastructural technology has concluded, there are many signs that the IT buildout is much closer to its end than its beginning" (Carr, 2003, p.10). The first and the third period exemplify the two outlined phases with the second period being transitional where continued construction of bulk elements like Content Delivery Networks (CDNs) and datacenters (though already improving the capability to distribute workloads dynamically) (Greenstein, 2019) started sharing the spotlight with mobility of access and application-oriented development of the ICT, for instance, the rise of Application Programming Interface (API) at the turn of millennia. The phase of establishment of the bulk part of ICT is captured in the literature on

Large Technical Systems (LTS) (Mayntz & Hughes, 1988) since an early prominent instantiation of ICT was telecommunication infrastructure (Davies, 1996). Later developments of ICT as control systems rooted in IT (Nightingale et al., 2003) already have a flavor of application-oriented mode. Beyond a systemic perspective of LTS theory, some studies consider specific ICT industries and mechanisms that grew atop the ICT infrastructure — Bresnahan et al. (2014) for mobile applications, Moore & Anderson (2012) for internet security, Jian et al. (2012) for the user-contributed production model of information goods — just to name few. In sum, in line with other studies, the results of the analysis indicate that the evolution of the ICT cluster is taking place and seems to have a direction towards agile, scalable, and omnipresent (mobile access) configuration with myriads of applications and devices. Thus, using the language of Helpman & Trajtenberg (1994), the next cycle of reaping the benefits of ICT development is on its way with the “new fruit” of ICT yet to be plucked.¹⁴

Moreover, if this path is to continue even further, more layers of functionality will be created making every next “frontend” layer more distant from the previous, lower level layers and providing more opportunities for forking. This calls for an inclusive process of integration, standardization, and compatibility preserving conditions for fair competition and accounting for societal welfare. The problem of achieving an inclusive consensus is already prominent in markets for ICT goods and services that are characterized by strong network effects (Shy, 2011) and grows larger due to the rise of platform business models (Belleflamme & Peitz, 2018); each firm has clear incentives to lock the network effects on itself, devising various strategies to cut-off, outpace or acquire competitors (Simcoe & Watson, 2019; Park et al., 2018; Cabral, 2018). Digital platforms and mega-apps like WeChat created by Tencent, GAFAM¹⁵, Baidu, Alibaba tech giants are examples, on the one hand, of success of these strategies but severely damage competition on the other hand (Prat & Valletti, 2019; Laitenberger, 2017). Thus, the current challenge for ICT development is to resolve the tension between the push for monopolization and pull for integration and compatibility (Gandal, 2002; Doganoglu & Wright, 2006) to achieve effective functioning within and among ICT applications and devices and yet preserve fair market conditions.

Pervasiveness of ICT classes. The estimation of the scope and scale of the economy’s reliance on a particular technology plays a role in the estimation of the pervasiveness of the technology under consideration. In turn, the question of pervasiveness or general applicability is one of the cornerstones of the General Purpose Technology (GPT) theory (Bresnahan & Trajtenberg, 1995). There are ongoing debates on the nature of pervasiveness and its measurement: using patent data (Hall & Trajtenberg, 2006; Feldman & Yoon, 2012; Graham & Iacopetta, 2014) or industrial diffusion patterns (Jovanovic & Rousseau, 2005; Castellacci, 2010). In this paper, I offer an alternative measure of pervasiveness based on the estimated scale and scope of reliance on ICT classes among industries derived from patent data. This goes in line with the view of Bekar et al. (2018) who suggest (i) to define GPTs “according to their micro-technological characteristics, not their macro-economic effects”, and (ii) to assess pervasiveness as one of such characteristics by distinguishing between cases when technology is (a) *widely used* and/or has (b) *many uses*. The former means that most of the economy relies on technology *at scale* even

¹⁴As opposed to the view of techno-pessimists such as Gordon (2016) and Cetto et al. (2016)

¹⁵Google, Amazon, Facebook, Apple, and Microsoft

if the technology has a single application, while the latter implies distinguished and multiple ways of using this technology or, in other words, a big *scope* of application. Thus, the calculated indicators of the scale and scope mirror this perspective on pervasiveness, and can contribute to the thread of literature on the empirical testing of technological pervasiveness by applying them to ICT classes. It is worth noting that in this paper I do not test whether either of the ICT classes is a GPT; instead, I estimate their pervasiveness in a novel way following the theoretical reasoning of Bekar et al. (2018).

As already mentioned, the industry–ICT matrices are in essence bipartite networks with industries and ICT classes being two different types of nodes. As explained in Section 3, the flow betweenness centrality (FBC) is used to proxy the scale of industry reliance on each ICT class. In application to the industry–ICT network, unweighted betweenness centrality would produce a similar picture compared to the calculation of the degree metric reflecting the number of incident industries. Instead, the FBC indicator takes the strength of the connection (i.e. weight of an edge) into account in the construction of the shortest path. Thus, FBC captures ICT classes that might be less frequently connected (in terms of the number of incident industries) but compensate for that by being intensely connected. This convenient property of the FBC indicator allows comparing it against the degree metric to derive conclusions about the scope and scale of reliance on each ICT class. In Figure 4, each observation is an ICT class (see Table 2) with the change of the degree and FBC indicators between the first and the third period as coordinates; the size of observations is defined by the magnitude of FBC of this ICT class in the third period.

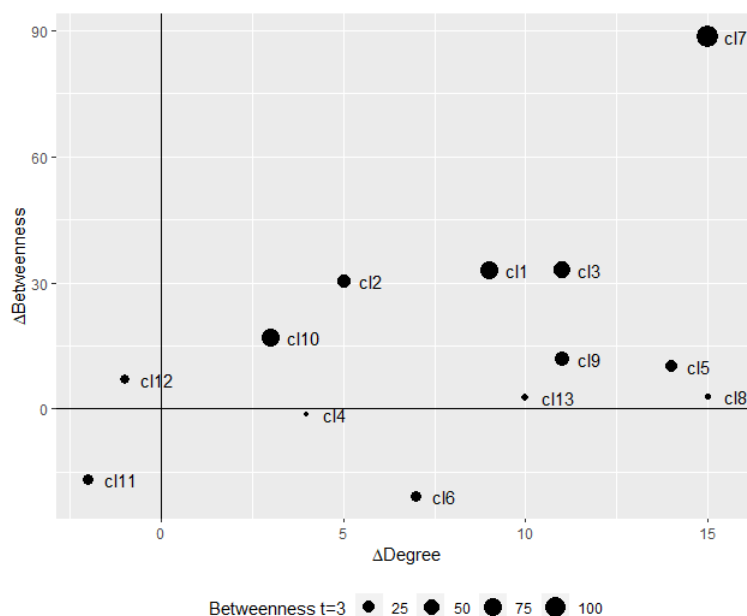


Figure 4: The change of scope (Δ degree) and scale (Δ FBC)

Δ Degree (x-axis) and Δ Betweenness (y-axis) are the first differences (change) of the respective metrics between the first and third periods. The size of the observations represents the absolute magnitude of the Flow Betweenness Centrality metric in the third period.

A comparison of Figure 4 and Figure 3 exposes the fact that the most central and connected ICT classes are not necessarily the most represented ones in the ICT cluster. Significantly smaller classes like Large–capacity information analysis (cl17) and Security (cl3) forged ahead of top–3

large classes High speed network (cl1), Information communication device (cl11), and Imaging and sound technology (cl10) in terms of increase in scope and scale, with class 7 acquiring a leading position with respect to both indicators and class 3 overcoming in scope and getting to parity in scale (see Appendix B). In general, the biggest ICT classes have a moderate number of incident industries but with stronger connections; by contrast, many smaller classes have numerous but weaker connections. The exceptions from this pattern are the two named classes 7 and 3.

Combining empirical estimations of scale and scope of diffusion and theoretical notions on pervasiveness, class 7 appears as the fittest candidate to be called pervasive by having many uses and being widely used. Classes 1, 3, and 10 are the closest competitors, though they significantly lag behind with respect to either scope or scale. An interesting dynamic unfolds for class 8 Artificial Intelligence (AI): it shows the largest increase in scope (the same as class 7) yet negligible growth in scale displaying an overall small absolute magnitude of the latter. Thus, class 8 AI has many applications but each of the applications doesn't rely at scale on AI. This finding of multiple but yet "shallow" applications is consistent with observations on AI diffusion (for example, as noted by [Bresnahan \(2019a\)](#) and [Brynjolfsson et al. \(2017\)](#)). A reasonable question arises: is AI pervasive? Turning again to [Bekar et al. \(2018\)](#) as the source of the proposed definition of the pervasiveness, the authors suggest that pervasive (in a GPT sense) can be called a technology that is rather widely used (at scale) than the one that has many uses (big scope); AI's pervasiveness can be overestimated because it is conflated with the pervasiveness of the whole ICT cluster it belongs to. Altogether, this conclusion raises a valid point of caution and can contribute to the discussion on whether AI is really a GPT ([Vannuccini & Prytkova, 2020](#)).¹⁶

In sum, the whole ICT cluster increases scope and scale of diffusion over four decades with discernible within-cluster differences among the constituting ICT classes.

Technological vs application relatedness. In this part I will discuss (i) obtained results on relatedness of knowledge and application bases among ICT technologies and (ii) analyse the change in industrial mix related to ICT.

Figure 5 is an empirical expression of the framework presented in Table 4; it plots technological and application relatedness as abscissa and ordinate respectively for each pair of ICT classes and contains two periods 1977–1990 ($t = 1$) and 2005–2020 ($t = 3$) to expose the dynamics of the indicators. Both indicators must be significant in either period for an ICT pair to be displayed in the graph.¹⁷ For 13 ICT classes, there are 78 pairs (excluding combination with itself) however only 22 have indicators that are significant, and, hence, present in Figure 5. Most of these pairs are located in the second quadrant, which implies similarity of application base among ICT technologies but specific knowledge base underlying each technology in the cluster. For example, based on Figure 5 one can observe that the following pairs of ICT technologies increase significantly the similarity of their application base: Information analysis and Security (pair 7–3), Human Interface and High Speed Computing (pair 9–5), High Speed Computing and Security (pair 5–3), and Imaging and Sound Technology and AI (pair 10–8). As suggested

¹⁶Proponent literature to this statement is represented by [Brynjolfsson et al. \(2019\)](#), [Trajtenberg \(2019\)](#), [Agrawal et al. \(2019a\)](#)

¹⁷Appendix C shows all 78 pairs.

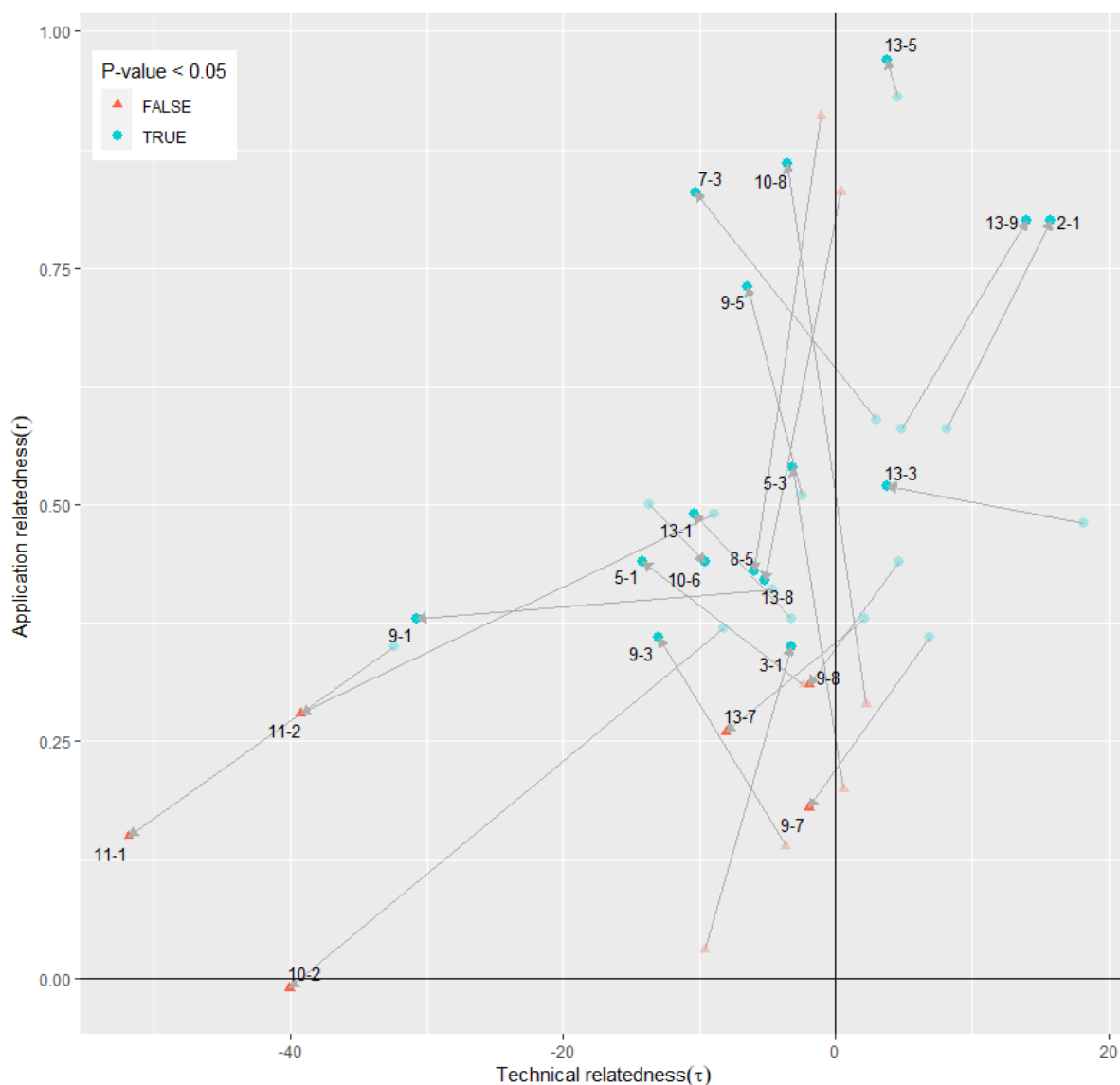


Figure 5: Movement of ICT pairs in relatedness space between periods 1 and 3

Two connected observations represent the same ICT pair in the first and the third period. Arrows originate in $t = 1$ position of the ICT pair going to its $t = 3$ position.

by the framework in Table 4, bundling and M&A patterns among firms, products and services embodying one of the listed ICT pair might be more pronounced in the industries related to these ICT pairs. To test the validity of the framework, further research is required to estimate bundling and M&A patterns in the industries related to ICT pairs from different quadrants and investigate their differences.

One exception from the overall picture is the pair formed by class 1 High speed network and class 2 Mobile communication that lies in the first quadrant. This means the pair not only exhibits similarity with respect to both relatedness indicators but strengthens its similarities over time. Overall, only 4 out of the 22 represented pairs increase their technological relatedness between the first and third period, while 12 out of 22 pairs experience growth of application relatedness. This indicates that already in the first period ICT technologies shared the application base and that tends to grow. As for the knowledge base, it remains delineated along the

ICT taxonomy classes. Perhaps, probing deeper levels of ICT classification, might reveal further proximity of knowledge base among more fine-grained, smaller ICT classes.

Completing the analytical exercise, in the next paragraphs I will focus on the AI ICT class to provide insights in its relatedness with other ICT classes and the change in application base AI has experienced. In addition to the estimations of pervasiveness, the position in the relatedness space and the structure of industry reliance on AI contributes to a better understanding of AI's development and to the construction of the broader context in which AI is considered.

AI and High speed computing. As can be seen in Figure 5, the two technologies (the pair of classes 8 and 5) exhibit a decrease in both indicators of relatedness. The pair remains in the second quadrant indicating similarity of the industrial base but distinct knowledge behind each technology. Between the two periods the relatedness of knowledge base decreases marginally, and the pair moves only slightly to the left. In contrast, application relatedness has clear dynamics plummeting from a strong to moderate level. The next paragraphs provide insights into this trend.

At the inception in the 1950s, AI was inseparable from *contemporary* powerful (high speed) computing machines. It was not yet a commercial technology but a scientific experiment, an attempt to simulate the higher functions of the human brain such as speaking and understanding language, sensory perception, reasoning, self-improvement or learning, forming abstractions, creative thinking, etc (McCarthy et al., 1955).¹⁸ This implies that in the past AI instantiations could exist if attached to an actor that has sufficiently powerful computing capacity and labor with programming skills. Within the 1977–1990 period, both resources were in their gestation and growth phases so initially scarce but growing rapidly. According to Beckhusen (2016), at the beginning of the 1970s in the US, “computers were large, expensive mainframes mostly used by governments, research laboratories, and manufacturing firms” while in 1990 already penetrated households and businesses. This also fueled the demand for IT workers whose number grew from 0.45 to 1.5 million between 1970 and 1990. These numbers include all types of IT workers with computer programmers alongside database administrators and computer network architects, hence only a fraction of which could be relevant to AI. For example, the enrollment in the introductory AI and Machine Learning courses in several US universities since 1990 is in the range of 200–300 up until approximately 2010 when the growth took off (see Appendix D). As suggested by Timmer & Van Ark (2005), investments in ICT capital and production of ICT goods in the EU lagged behind the US, therefore a similar dynamics could unfold in the EU with the whole timeline shifted in time.

In fact, the relation between AI production and the possession of resources for that was the opposite: at the pre-commercial phase of AI the possession of computing capacity and programming skills didn't stem from the incentive to develop AI but eventually allowed doing so, as opposed to the current substantially commercial phase when the goal of AI development and/or usage consequently drives the decision to acquire the necessary resources. With the invention and gradual growth of commercial AI forms, the markets for hardware and labor with programming skills started experiencing a positive shock. This triggered an ameliorative loop

¹⁸Mohamed et al. (2020) summarize the evolution of AI from its onset until now claiming that “AI has seen itself elevated from an obscure domain of computer science into technological artefacts embedded within and scrutinised by governments, industry and civil society”.

among markets for AI-related labor, hardware and AI solutions (software). The form of provision of labor and hardware resources to developers and adopters of AI solutions is one of these two: production in-house or purchase (make-or-buy dilemma). In particular, over the decades computing power became not only more affordable (Flamm, 2019) but also more accessible facilitated by cloud computing (Byrne et al., 2018). The growing quality and availability of both AI solutions and AI-related resources facilitates AI adoption in industries where the commercial value is either not obvious (experimentation) or requires additional adjustments (implementation lags) of production processes, capital, business models, etc. Thus, those novel industries that adopt AI are likely not the ones that relied before on high speed computing for their tasks and functions. Moreover, high speed computing is bound to AI in industries that build functionality and maintain AI systems in-house and have not purchased access to it from AI provider with subsequent servicing.

Looking at the empirical results, the most recent industries that appeared in Table 6 such as pharmaceutical (e.g. drug discovery, medical imaging), scientific R&D (AI as Invention of Method of Invention (IMI) (Cockburn et al., 2019)), employment (e.g. HR algorithms), security and investigation (e.g. predictive policing, suspect identification with visual recognition) are rather industries buying AI solutions for their data processing. As noted earlier, the first production of AI systems emerged atop of the possession of computing and programming labor resources hence producers of AI must be the long-standing holders of both resources. In terms of the current analysis, potential AI-producing industries are likely to be (i) listed among industries related to both AI and High speed computing since the first period (incumbent industries) and (ii) ranked high in terms of strength of connection with both ICT classes (by the sum of ranks over three periods).

| | ISIC rev. 4 | Class 5: HSC | Class 8: AI |
|----|---|--------------|-------------|
| 59 | Motion picture, video, sound recording, music publishing activities | - | 1 |
| 62 | Computer programming, consultancy and related activities | 1 | 2 |
| 63 | Information service activities | 2 | 3 |
| 26 | Manufacturing of computer, electronic and optical products | 3 | - |

Table 5: Top-3 positions in the intertemporal ranking of industries connected to AI and HSC

Not surprisingly, two industries that fulfill both criteria are 62-Computer programming, consultancy and related activities and 63-Information service activities. As shown in Table 5, these two incumbent industries occupy the second and third positions in the intertemporal ranking for the strength of connection with the AI class and first and second for the High speed computing. The production process of digital products and services in these activities involves precisely the two mentioned production factors — labor with programming skills and computing machinery as capital. An interesting finding is which industry completes the top-3 ranking for each of the ICT classes. Concerning AI, the first position in the ranking is held by the endemic to AI industry 59-Motion picture, video, sound recording, music publishing activities. As for High speed computing, the third position in the ranking belongs to industry 26-Manufacturing of computer, electronic and optical products. Altogether, this breaks down AI technology into its basic components: perception through sensory data, information-processing algorithms, and computing machinery, echoing the composition suggested by Taddy (2019).

In sum, the number of AI-related industries grew apparently due to the increase in the number of AI-using industries. These industries adopt AI solutions developed and managed by an AI-provider who bundles the provision of AI algorithms with computing power required to support the application. The joint presence of AI and High Speed Computing becomes localized and limited to a small number of AI-producing industries (i.e. 62 and 63), while AI-using industries show connection to AI only. This results in a decreasing application relatedness between class 8 AI and class 5 High speed computing.

| 1977-1990 t=1 | 1991-2004 t=2 | 2005-2020 t=3 | Code | ISIC rev.4 Description |
|------------------|------------------|------------------|------|---|
| | | | 18 | Printing and reproduction of recorded media |
| | | | 21 | Manufacture of basic pharmaceutical products and pharmaceutical preparations |
| | | | 26 | Manufacture of computer, electronic and optical products |
| | | | 32 | Other manufacturing |
| | | | 51 | Air transport |
| | | | 58 | Publishing activities |
| | | | 59 | Motion picture, video, sound recording, music publishing activities |
| | | | 61 | Telecommunications |
| | | | 62 | Computer programming, consultancy and related activities |
| | | | 63 | Information service activities |
| | | | 64 | Financial service activities, except insurance and pension funding |
| | | | 65 | Insurance, reinsurance and pension funding, except compulsory social security |
| | | | 66 | Activities auxiliary to financial service and insurance activities |
| | | | 69 | Legal and accounting activities |
| | | | 71 | Architectural and engineering activities; technical testing and analysis |
| | | | 72 | Scientific research and development |
| | | | 73 | Advertising and market research |
| | | | 74 | Other professional, scientific and technical activities |
| | | | 78 | Employment activities |
| | | | 80 | Security and investigation activities |
| | | | 82 | Office administrative, office support and other business support activities |
| | | | 86 | Human health activities |
| | | | 87 | Residential care activities |
| | | | 90 | Creative, arts and entertainment activities |
| | | | 91 | Libraries, archives, museums and other cultural activities |

Table 6: Dynamics of industry reliance on AI

AI and Imaging and sound technology. The pair (class 8 and class 10 respectively) migrates from the first to the second quadrant by decreasing its technological relatedness. Similar to the previous pair of AI with High speed computing, this pair exhibits a significant change in the application relatedness but in this case it surges to a strong level.

As it has been pointed out in the previous paragraphs, one of the onset goals of AI is to replicate the complex function of sensory perception in different modalities (visual, audio, tactile). Understanding the algorithms behind the processing of *unstructured*, raw sensory data that results in structured information (pattern recognition) has been a challenge due to algorithms' non-deterministic nature; only after studying the same processes in living organisms the idea of artificial neurons and their "assembly formation", i.e neural networks capable of calculation have been formulated.¹⁹ Even after proof of concept in the late 1950s and 1960s, it took almost five decades to overcome initial limitations and find the way to scale up a simple perceptron to a modern version of Artificial Neural Networks (ANN) that surpass human performance in a set of tasks (see Eckersley et al. (2017)). It turned out that pattern recognition capabilities of ANNs that could harness raw sensory data without preprocessing can also work as well with structured, e.g. trading data, demographic data, medical records, and fuzzy-structured, e.g. lan-

¹⁹Ground work by McCulloch & Pitts (1943) and Hebb (2005)

guage, clickstreams, data. Thus, starting in the domain of pure logic and deterministic rules with the symbolic approach, eventually, AI has evolved eventually into a non-deterministic, highly perceptive instantiation, representing the connectionist approach. The ability to transform vast amounts of structured and especially unstructured data into information let connectionist AI in many different industries where such information (in a form of inference and/or prediction) has a value (Agrawal et al., 2019c). Thus, Image and sound technology became an ultimate tool of data collection for further processing by AI systems, which is reflected in the surging growth of application relatedness between these two ICT classes.

5 Conclusion

The importance of ICT for the functioning of any economic system cannot be underestimated. However, studies on the impact of ICT often considered this cluster as a monolith block. In this paper, I distinguish a set of ICT technologies employing the new ICT taxonomy from OECD and PATENTSCOPE AI Index and estimate ICTs' connections with industries using the Algorithmic Links with Probabilities method. The construction of a fine-grained industry-technology map allows assessing the structure and evolution of industry reliance on ICT. This required the application of several text analysis techniques to break down industrial descriptions into sets of keywords to match them with patents' abstracts and titles. The subsequent application of network analysis and relatedness indicators helps to uncover patterns and regularities in the structure and dynamics of the constructed ICT technology-industry network.

The results indicate that the ICT cluster shows signs of a "phase transition", passing the phase of building bulk elements of the infrastructure and around the 2000s entering the phase of working on the functionality for business applications deployment and users' convenience. More application-oriented technologies like mobile communication, information analysis, security and human interface show significant and persistent growth of their shares in the ICT cluster in the EU28 region. In contrast, more mature technologies that represent major physical components of the infrastructure such as high speed network and information communication device recede though keeping a strong presence in the cluster. The inclusion of the industries into consideration allows looking into the structure of connections between ICT technologies and industries and its dynamics. Despite being the largest in the ICT system, high speed network, information communication device and imaging and sound technology are not the most central and connected ones. Instead, information analysis is forging far ahead with regard to both scale and scope, penetrating an increasing number of industries and strengthening its industrial connections. Said differently, information analysis moves rapidly towards the center of the knowledge base of the ICT-related industries compared to other ICT technologies in the cluster. Security technology occupies the second position after information analysis by overcoming the largest ICT classes in scope and getting to parity in scale. Overall, the biggest ICT classes have a moderate number of incident industries but with stronger connections; by contrast, many smaller classes (except for information analysis and security) have numerous but weaker connections.

According to the framework represented in Table 4, the position of the overwhelming majority of ICT pairs the second quadrant indicates shared industrial base though distinct knowledge underlying each ICT technology. Some pairs strengthen their position in the second quadrant

such as: Information analysis and Security, Human Interface and High Speed Computing, High Speed Computing and Security, and Imaging and Sound Technology and AI. Following the suggested framework in Table 4, for the pairs in the second quadrant the visibility of commercial value might be a factor that navigates the innovation process. Pursuit of commercial value might lead to more pronounced and systematic bundling of products and services and/or M&A among firms producing goods embodying these ICT technologies than for ICT located elsewhere in the relatedness space. Overall, among ICT pairs, application relatedness tends to increase over time (12 pairs out of 22) while technological relatedness appears rather stable (4 pairs out of 22).

A special focus of the analysis is placed on AI technology among the ICT cluster. On the one hand, AI is a novel, fast-growing technology that enters the commercial phase and is subject to intensive development and experimentation. Multiple applications, unprecedented potential for automation and billions in generated revenues make AI a fruitful topic to study. On the other hand, the research on AI might benefit from putting this technology into context, and study AI in relation and in comparison with other technologies. In this paper, ICT classes serve simultaneously as potential complementors and as benchmark, building a framework for AI's evaluation. An interesting finding concerning AI is that it shows the largest increase in scope yet negligible growth in scale and its small absolute magnitude. This points at multiple but yet "shallow" connections between AI and industries going in line with AI's gestation and early growth phase. In the technical literature, it is reported that both AI algorithms and hardware represent ad hoc solutions that lack flexibility (Sze et al., 2020; Hooker, 2020), and producers are only at the beginning of addressing this issue. In particular, the identified connection between High Speed Computing and AI technologies shows decreasing but significant application relatedness perhaps due to the acquisition of AI-using industries that largely rely on computing power in the cloud or adopt pretrained AI models and themselves do not employ powerful hardware. This reflects the actual dilemma between fragmentation into specialized hardware and integration of broad functionality under a platform chip that the semiconductor industry is currently facing (Prytkova & Vannuccini, 2020). If the semiconductor industry will decide in favor of specialized hardware (appealing to AI producers) the trend of purchasing access to AI solutions run by AI producers among industries is likely to continue; the application relatedness between High Speed Computing and AI might experience a further decrease. This would also mean the completion of AI upstreaming, with AI becoming a fully-fledged, distinct industry. In addition to the described challenges to be resolved at the supply side, the reorganization of production and business models on the demand side and adoption lags are likely to delay AI deployment at scale as well.

A further disaggregation of the ICT cluster into more fine-grained technologies might improve the precision of estimation of technological relatedness among the technologies in the cluster. The framework in Table 4 has to be tested, hence further research is required to estimate bundling and M&A patterns in the industries related to ICT pairs from different quadrants and investigate their differences. An investigation into the identified connections between ICT technologies akin to the ones involving AI discussed in this paper can be not only an interesting exercise for historians of technology but can also inform economists studying technological diffusion, system products and network externalities, and policy-makers in identifying

related markets and technologies even across industrial boundaries. As for the estimation of the industry–technology connections, the inherent shortcoming lies in the organization of industrial classification based on the predominant activity; hence in the description the related activities can be underrepresented. However, related activities can be picked up at least partially if their keywords are mentioned together with keywords of the predominant activities in a patent document. Thus, the keyword–based matching partially tackles the problem of false positives by placing no restriction on the matching. The same advantage might turn into a disadvantage when it creates non–existing connections (false negatives). The bigrams and keywords refinement (expansion or/end replacement) with synonyms has been used to reduce the number of false negatives.

References

- Agrawal, A., Gans, J., & Goldfarb, A. (2019a). Economic policy for artificial intelligence. *Innovation Policy and the Economy*, 19(1), 139–159.
- Agrawal, A., Gans, J., & Goldfarb, A. (2019b). *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019c). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2), 31–50.
- Anderson, S. (2012). Advertising on the internet. *The Oxford handbook of the digital economy*, (pp. 355–396).
- Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: have computers changed the labor market? *The Quarterly journal of economics*, 113(4), 1169–1213.
- Bakos, Y. & Brynjolfsson, E. (2001). Aggregation and disaggregation of information goods: Implications for bundling, site licensing, and micropayment systems. In *Lectures in E-Commerce* (pp. 103–122). Springer.
- Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the european union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9), 1252–1268.
- Basu, S. & Fernald, J. (2007). Information and communications technology as a general-purpose technology: Evidence from us industry data. *German Economic Review*, 8(2), 146–173.
- Basu, S., Fernald, J. G., Oulton, N., & Srinivasan, S. (2003). The case of the missing productivity growth, or does information technology explain why productivity accelerated in the united states but not in the united kingdom? *NBER macroeconomics annual*, 18, 9–63.
- Beckhusen, J. (2016). *Occupations in information technology*. Technical report, US Department of Commerce, Economics and Statistics Administration, US Census Bureau.
- Bekar, C., Carlaw, K., & Lipsey, R. (2018). General purpose technologies in theory, application and controversy: a review. *Journal of Evolutionary Economics*, 28(5), 1005–1033.
- Belleflamme, P. & Peitz, M. (2018). Platforms and network effects. In *Handbook of Game Theory and Industrial Organization, Volume II*. Edward Elgar Publishing.
- Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1), 69–87.
- Bresnahan, T. (2019a). Artificial intelligence technologies and aggregate growth prospects.
- Bresnahan, T. & Yin, P.-L. (2017). Adoption of new information and communications technologies in the workplace today. *Innovation policy and the economy*, 17(1), 95–124.

- Bresnahan, T. F. (2019b). Technological change in ict in light of ideas first learned about the machine tool industry. *Industrial and Corporate Change*, 28(2), 331–349.
- Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics*, 117(1), 339–376.
- Bresnahan, T. F., Davis, J. P., & Yin, P.-L. (2014). Economic value creation in mobile applications. In *The changing frontier: Rethinking science and innovation policy* (pp. 233–286). University of Chicago Press.
- Bresnahan, T. F. & Trajtenberg, M. (1995). General purpose technologies ‘engines of growth’? *Journal of econometrics*, 65(1), 83–108.
- Brynjolfsson, E. & Hitt, L. (1995). Information technology as a factor of production: The role of differences among firms. *Economics of Innovation and New technology*, 3(3-4), 183–200.
- Brynjolfsson, E. & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic perspectives*, 14(4), 23–48.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics*. Technical report, National Bureau of Economic Research.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial intelligence and the modern productivity paradox. *The Economics of Artificial Intelligence: An Agenda*, (pp.23).
- Byrne, D., Corrado, C., & Sichel, D. E. (2018). *The rise of cloud computing: minding your P’s, Q’s and K’s*. Technical report, National Bureau of Economic Research.
- Cabral, L. (2018). Standing on the shoulders of dwarfs: Dominant firms and innovation incentives.
- Carr, N. G. (2003). It doesn’t matter. *Educause Review*, 38, 24–38.
- Castellacci, F. (2008). Technological paradigms, regimes and trajectories: Manufacturing and service industries in a new taxonomy of sectoral patterns of innovation. *Research Policy*, 37(6-7), 978–994.
- Castellacci, F. (2010). Structural change and the growth of industrial sectors: Empirical test of a gpt model. *Review of income and wealth*, 56(3), 449–482.
- Cette, G., Fernald, J., & Mojon, B. (2016). The pre-great recession slowdown in productivity. *European Economic Review*, 88, 3–20.
- Church, J. & Gandal, N. (2005). Platform competition in telecommunications. In *The Handbook of Telecommunications*, volume 2 (pp. 119–155). Elsevier.

- Cockburn, I. M., Henderson, R., & Stern, S. (2019). The impact of artificial intelligence on innovation. *The Economics of Artificial Intelligence: An Agenda*, (pp. 115).
- Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and us economic growth. *Review of income and wealth*, 55(3), 661–685.
- Corrado, C., Lengermann, P., Beaulieu, J. J., & Bartelsman, E. J. (2007). Sectoral productivity in the united states: Recent developments and the role of it. *German Economic Review*, 8(2), 188–210.
- David, P. A. (2007). Path dependence: a foundational concept for historical social science. *Econometrica*, 1(2), 91–114.
- Davies, A. (1996). Innovation in large technical systems: the case of telecommunications. *Industrial and Corporate Change*, 5(4), 1143–1180.
- Dewan, S. & Min, C.-k. (1997). The substitution of information technology for other factors of production: A firm level analysis. *Management science*, 43(12), 1660–1675.
- Diewert, W. E. & Fox, K. J. (1999). Can measurement error explain the productivity paradox? *The Canadian Journal of Economics/Revue canadienne d'Economique*, 32(2), 251–280.
- Doganoglu, T. & Wright, J. (2006). Multihoming and compatibility. *International Journal of Industrial Organization*, 24(1), 45–67.
- Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3), 147–162.
- Eckersley, P., Nasser, Y., et al. (2017). Eff ai progress measurement project.
- Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in us metropolitan areas. *Regional Studies*, 49(5), 752–766.
- Federico, G., Morton, F. S., & Shapiro, C. (2020). Antitrust and innovation: Welcoming and protecting disruption. *Innovation Policy and the Economy*, 20(1), 125–190.
- Feldman, M. P. & Yoon, J. W. (2012). An empirical test for general purpose technology: an examination of the cohen–boyer rdna technology. *Industrial and Corporate Change*, 21(2), 249–275.
- Fernald, J. G. (2015). Productivity and potential output before, during, and after the great recession. *NBER macroeconomics annual*, 29(1), 1–51.
- Fitchard, K. (2003). Crossing over: The journey to packets. *Telephony*, 244(16), 40–40.
- Flamm, K. (2019). Measuring moore’s law: evidence from price, cost, and quality indexes. In *Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.
- Freeman, C. (1994). The economics of technical change. *Cambridge journal of economics*, 18(5), 463–514.

- Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies*, 41(5), 685–697.
- Gandal, N. (2002). Compatibility, standardization, and network effects: Some policy implications. *Oxford Review of Economic Policy*, 18(1), 80–91.
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–74.
- Gonzalez, J. & Ponemon, L. (2019). *2019 Intangible Assets Financial Statement Impact Comparison Report*. Technical report, Ponemon Institute LLC.
- Gordon, R. J. (2016). Perspectives on the rise and fall of american growth. *Education*, 2(2.5), 3.
- Graf, H. (2012). Inventor networks in emerging key technologies: information technology vs. semiconductors. *Journal of Evolutionary Economics*, 3(22), 459–480.
- Graham, S. J. & Iacopetta, M. (2014). Nanotechnology and the emergence of a general purpose technology. *Annals of Economics and Statistics/Annales D'Économie et de Statistique*, (115/116), 25–55.
- Greenstein, S. (2019). Digital infrastructure. In *Economics of Infrastructure Investment*. University of Chicago Press.
- Greenstein, S. M. & Spiller, P. T. (1996). *Estimating the welfare effects of digital infrastructure*. Technical report, National Bureau of Economic Research.
- Guerrieri, P. & Padoan, P. C. (2007). *Modelling ICT as a general purpose technology*. The College of Europe.
- Hall, B. H. & Trajtenberg, M. (2006). 14. uncovering general purpose technologies with patent data1. *New Frontiers in the Economics of Innovation and New Technology: Essays in Honour of Paul A. David*, (pp. 389).
- Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- Helpman, E. & Trajtenberg, M. (1994). *A time to sow and a time to reap: growth based on general purpose technologies*. Technical report, National Bureau of Economic Research.
- Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., et al. (2018). The principle of relatedness. In *International conference on complex systems* (pp. 451–457).: Springer.
- Hidalgo, C. A. & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570–10575.
- Hooker, S. (2020). The hardware lottery. *arXiv preprint arXiv:2009.06489*.

- Inaba, T. & Squicciarini, M. (2017). Ict: A new taxonomy based on the international patent classification. *OECD iLibrary*.
- Jian, L., MacKie-Mason, J., Chiao, B., Levchenko, A., Zellner, A., Kmenta, J., Dreze, J., & Oberhofer, W. (2012). Incentive-centered design for user-contributed content. *The Oxford Handbook of the Digital Economy*, (pp. 399).
- Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2003). Growth of us industries and investments in information technology and higher education. *Economic Systems Research*, 15(3), 279–325.
- Jovanovic, B. & Rousseau, P. L. (2005). General purpose technologies. In *Handbook of economic growth*, volume 1 (pp. 1181–1224). Elsevier.
- Kortum, S. & Putnam, J. (1997). Assigning patents to industries: tests of the yale technology concordance. *Economic Systems Research*, 9(2), 161–176.
- Laitenberger, J. (2017). Eu competition law in innovation and digital markets: fairness and the consumer welfare perspective. In *MLex / Hogan Lovells event*.
- Lybbert, T. J. & Zolas, N. J. (2014). Getting patents and economic data to speak to each other: An “algorithmic links with probabilities” approach for joint analyses of patenting and economic activity. *Research Policy*, 43(3), 530–542.
- Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, 31(2), 247–264.
- Malerba, F. & Orsenigo, L. (1997). Technological regimes and sectoral patterns of innovative activities. *Industrial and corporate change*, 6(1), 83–118.
- Maskell, P. & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge journal of economics*, 23(2), 167–185.
- Mayntz, R. & Hughes, T. (1988). *The Evolution of Large Technical Systems*. Campus.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). Proposal for the 1956 dartmouth summer research project on artificial intelligence, dartmouth college.
- McCulloch, W. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Michaels, G., Natraj, A., & Van Reenen, J. (2014). Has ict polarized skill demand? evidence from eleven countries over twenty-five years. *Review of Economics and Statistics*, 96(1), 60–77.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial ai: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684.
- Moore, T. & Anderson, R. (2012). Internet security. In *The Oxford Handbook of the Digital Economy* (pp. 572–99). Oxford University Press Oxford.
- Morton, F. S. & Dinielli, D. (2020). *Roadmap for an Antitrust Case Against Facebook*. Technical report, Omidyar Network.

- Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, 94(1), 23–48.
- Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic geography*, 87(3), 237–265.
- Nightingale, P., Brady, T., Davies, A., & Hall, J. (2003). Capacity utilization revisited: software, control and the growth of large technical systems. *Industrial and Corporate Change*, 12(3), 477–517.
- Park, K., Seamans, R., & Zhu, F. (2018). Multi-homing and platform strategies: Historical evidence from the us newspaper industry. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (18-032).
- Perez, C. (2010). Technological revolutions and techno-economic paradigms. *Cambridge journal of economics*, 34(1), 185–202.
- Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *The AI Index 2019 Annual Report*. Technical report, AI Index Steering Committee, Human-Centered AI Institute, Stanford University.
- Prat, A. & Valletti, T. M. (2019). Attention oligopoly. *Available at SSRN 3197930*.
- Prytkova, E. & Vannuccini, S. (2020). On the basis of brain: Neural-network-inspired change in general purpose chips.
- Roach, S. S. (1987). America's technology dilemma: A profile of the information economy: Morgan stanley special economic study. *Morgan Stanley*.
- Rosenberg, N. (1963). Technological change in the machine tool industry, 1840-1910. *Journal of economic history*, (pp. 414–443).
- Rosenberg, N. (1972). Factors affecting the diffusion of technology. *Explorations in economic history*, 10(1), 3.
- Schmoch, U., Laville, F., Patel, P., & Frietsch, R. (2003). Linking technology areas to industrial sectors. *Final Report to the European Commission, DG Research*.
- Shy, O. (2011). A short survey of network economics. *Review of Industrial Organization*, 38(2), 119–149.
- Simcoe, T. & Watson, J. (2019). Forking, fragmentation, and splintering. *Strategy Science*, 4(4), 283–297.
- Solow, R. M. (1987). New york times book review. *July*, 12(1987), 36.
- Steinmueller, W. E. (1992). The economics of flexible integrated circuit manufacturing technology. *Review of Industrial Organization*, 7(3-4), 327–349.

- Steinmueller, W. E. (1996). Technological infrastructure in information technology industries. In *Technological Infrastructure Policy* (pp. 117–139). Springer.
- Steinmueller, W. E. (2002). Knowledge-based economies and information and communication technologies. *International Social Science Journal*, 54(171), 141–153.
- Syverson, C. (2017). Challenges to mismeasurement explanations for the us productivity slowdown. *Journal of Economic Perspectives*, 31(2), 165–86.
- Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2020). How to evaluate deep neural network processors: Tops/w (alone) considered harmful. *IEEE Solid-State Circuits Magazine*, 12(3), 28–41.
- Taddy, M. (2019). The technological elements of artificial intelligence. *The Economics of Artificial Intelligence: An Agenda*, (pp. 61).
- Takahashi, T. & Namiki, F. (2003). Three attempts at “de-wintelization”: Japan’s tron project, the us government’s suits against intel, and the entry of java and linux. *Research Policy*, 32(9), 1589–1606.
- Timmer, M. P. & Van Ark, B. (2005). Does information and communication technology drive eu-us productivity growth differentials? *Oxford Economic Papers*, 57(4), 693–716.
- Trajtenberg, M. (2019). Artificial intelligence as the next gpt. *The Economics of Artificial Intelligence: An Agenda*, (pp. 175).
- UN (2008). *International Standard Industrial Classification of All Economic Activities Revision 4 (ISIC rev.4)*. Number 4. United Nations Publications.
- Van Ark, B., Inklaar, R., & McGuckin, R. H. (2003). Ict and productivity in europe and the united states where do the differences come from? *CESifo Economic Studies*, 49(3), 295–318.
- Vannuccini, S. & Prytkova, E. (2020). Artificial intelligence’s new clothes? from general purpose technology to large technical system. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3704011>.
- WIPO (2019). *PATENTSCOPE Artificial Intelligence Index*. Technical report, WIPO.

Appendices

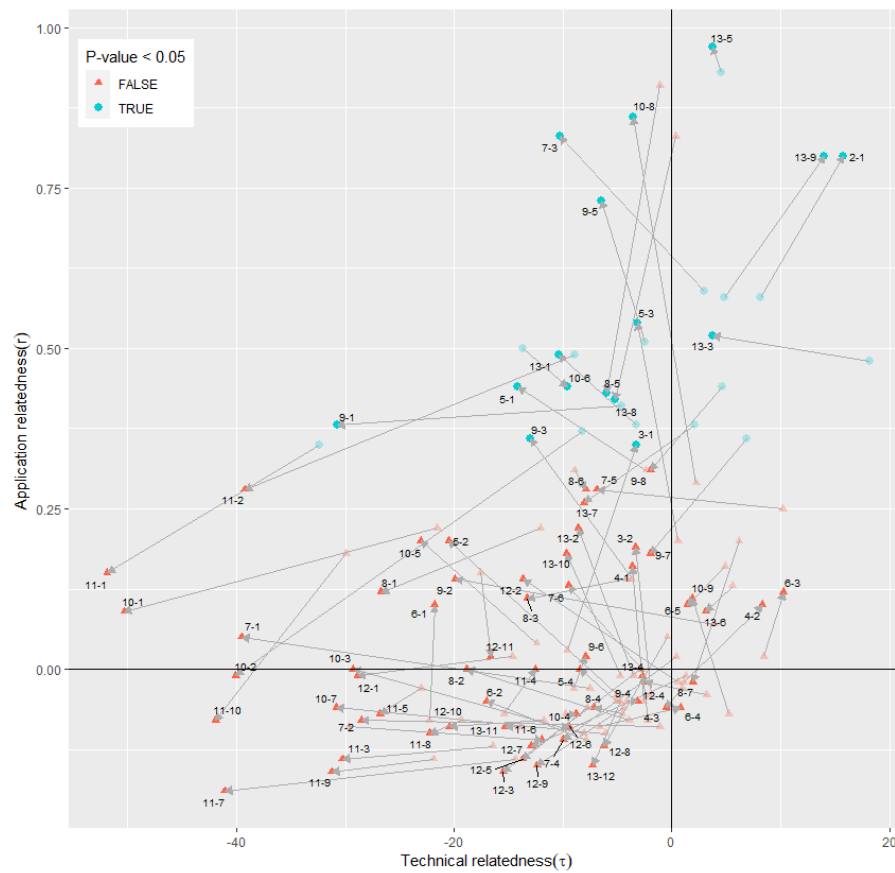
Appendix A Descriptive statistics of industry–ICT bipartite network

| Period | | 1977-1990 | 1991-2004 | 2005-2020 |
|-----------|----------|----------------|-----------|-----------|
| | | Degree | | |
| ICT | <i>n</i> | 13 | 13 | 13 |
| | mean | 8.62 | 13.92 | 16.38 |
| | median | 9.00 | 14.00 | 20.00 |
| | min | 4 | 1 | 3 |
| | max | 14 | 22 | 25 |
| ISIC | <i>m</i> | 26 | 32 | 36 |
| | mean | 4.31 | 5.66 | 5.92 |
| | median | 3.00 | 6.50 | 7.50 |
| | min | 1 | 1 | 1 |
| | max | 13 | 12 | 12 |
| | | Density | | |
| Bipartite | | 0.33 | 0.44 | 0.46 |

Appendix B Flow betweenness centrality and degree indicators

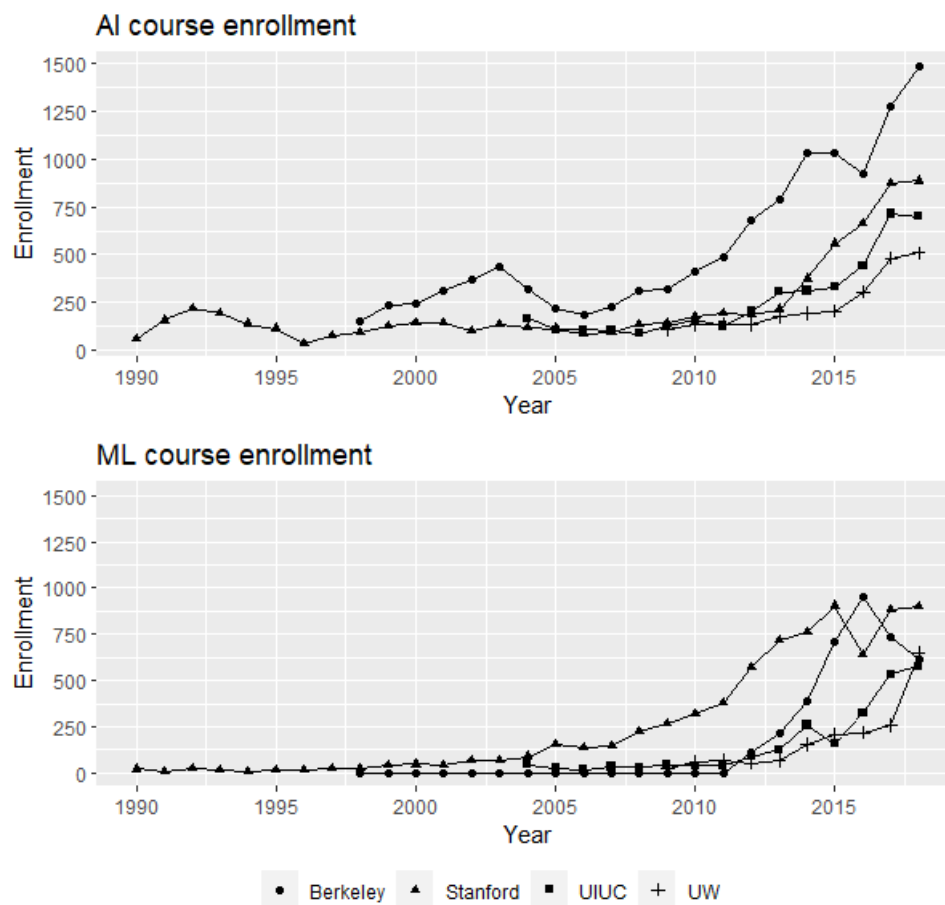
| Period ICT class | t=1 | | t=2 | | t=3 | |
|---------------------|--------|-------|--------|-------|--------|--------|
| | Degree | FBC | Degree | FBC | Degree | FBC |
| cl1 | 7 | 33.71 | 14 | 70.88 | 16 | 66.78 |
| cl2 | 4 | 3.27 | 8 | 20.19 | 9 | 33.60 |
| cl3 | 14 | 22.48 | 21 | 44.70 | 25 | 55.46 |
| cl4 | 6 | 4.82 | 9 | 2.41 | 10 | 3.50 |
| cl5 | 6 | 12.92 | 14 | 20.54 | 20 | 22.95 |
| cl6 | 13 | 37.70 | 16 | 27.98 | 20 | 17.04 |
| cl7 | 9 | 20.82 | 20 | 76.17 | 24 | 109.37 |
| cl8 | 9 | 1.99 | 20 | 5.82 | 24 | 4.84 |
| cl9 | 13 | 24.89 | 22 | 30.18 | 24 | 36.84 |
| cl10 | 9 | 53.07 | 13 | 61.24 | 12 | 70.15 |
| cl11 | 8 | 33.58 | 6 | 17.64 | 6 | 16.82 |
| cl12 | 4 | 7.61 | 1 | 0.00 | 3 | 14.49 |
| cl13 | 10 | 4.30 | 17 | 7.44 | 20 | 6.96 |

Appendix C Technological and application relatedness



Note: Arrows originate from $t = 1$ going to $t = 3$ position

Appendix D Enrollment in AI and ML courses, US universities



Note: Reproduced from the AI Index report 2019 (Perrault et al., 2019)

IMPRESSUM

Jena Economic Research Papers

ISSN 1864-7057

Friedrich Schiller University Jena

Faculty of Economics and Business Administration

Carl-Zeiss-Str. 3

D-07743 Jena, Germany

Email: office.jerp@uni-jena.de

Editor: Silke Übelmesser

Website: www.wiwi.uni-jena.de/en/jerp

© by the author