Papers in Evolutionary Economic Geography

19.18

Using structural diversity to measure the complexity of technologies

Tom Broekel

Utrecht University Human Geography and Planning

USING STRUCTURAL DIVERSITY TO MEASURE THE COMPLEXITY OF TECHNOLOGIES

Tom Broekel*
Department of Urban Geography and Spatial Planning
Utrecht University
Princetonlaan 8A, 3584 CB Utrecht, The Netherlands
t.broekel@uu.nl

Published in PLoS ONE, DOI:10.1371/journal.pone.0216856

May 22, 2019

ABSTRACT

The paper introduces structural diversity as a new approach to quantify the complexity of technologies. By modeling technologies as combinatorial networks, a measure of technological complexity is derived that represents the diversity of (sub-)network topologies in these networks. It is further argued that this measure can be empirically approximated with the Network Diversity Score (NDS).

The paper also presents an application of this approach to European patent data from 1980 to 2015. On this basis, the measure of structural diversity is shown to replicate a number of stylized facts commonly associated with technological complexity: Complexity increases over time and younger technologies are more complex than older technologies. Complex technologies are also associated to larger R&D efforts and require more collaborative R&D activities. Lastly, when controlling for technologies' size, technologies scoring high on structural diversity are also shown to concentrate in space.

Keywords Complexity, technology, patents, technological complexity, network, diversity

JEL: O11, O31, N70

1 Introduction

1.1 Measuring the complexity of knowledge

The complexity of knowledge is seen as a crucial explanatory dimension of technological development and economic success (Romer, 1990; Dalmazzo, 2002; Sorenson, 2005; Hidalgo and Hausmann, 2009; Balland and Rigby, 2017). The higher difficulty of inventing and learning complex knowledge is argued to require larger economic efforts of entering these domains. This hinders the diffusion of such knowledge among economic agents (Kogut and Zander, 1992; Sorenson et al., 2006). Consequently, complex knowledge can be expected to be more exclusive and therefore to possess more economic value (Simon, 1962; Rivkin, 2000).

*E-mail: tombroekel@me.com, Web: tombroekel.de

However, empirical studies analyzing technological and innovation processes frequently rely on simple counts of knowledge inputs or outputs (e.g., patents, number of new products), and thereby fail to capture this dimension of knowledge (Balland and Rigby, 2017). This shortcoming seems to be less a matter of recognizing the dimension's importance and more of a lack of a convincing (quantitative) measure of knowledge complexity (Balland and Rigby, 2017). There have been many attempts in different disciplines to measure knowledge complexity (Goldreich et al., 1998; Goldreich and Petrank, 1999; Ye, 2017; Fleming and Sorenson, 2001; Pintea and Thompson, 2007; Balland and Rigby, 2017; Kim and Anand, 2018). However, for most of these, it is still unknown if and if so how they are applicable to real world data as well as whether they allow for differentiating knowledge domains according to degrees of complexity.

Two notable exceptions in this respect are the works of Fleming and Sorenson (2001) and Balland and Rigby (2017). Based on an N/K model, *Fleming and Sorenson* quantify the degree of interdependence inherent to subcomponents of a knowledge domain, which these authors interpret as an approximation of knowledge complexity. The applicability of this approach to patent data is demonstrated in multiple empirical studies (Sorenson and Fleming, 2004; Sorenson et al., 2006).[2] In contrast, Balland and Rigby (2017) apply *Hidalgo and Hausmmann's economic complexity index*, which was originally designed to assess the complexity of countries' export and employment patterns, to patent data, and thereby obtain an index of knowledge complexity (Hidalgo and Hausmann, 2009; Balland and Rigby, 2017).

However, both approaches rely on strong assumptions. The complexity measure of *Fleming and Sorenson* is build on the idea that subcomponents of complex inventions are difficult to combine, which translates into such combinations being relatively infrequent (Fleming and Sorenson, 2001). Yet, economic reasons unrelated to difficulties in inventive processes might influence the frequency of knowledge combinations. For instance, the lack of some combinations' market potential can result in minimal attention from researchers. Alternatively, as noted in other works of these authors, it may instead be the range of applications shaping the combinatorial frequency, which may or may not reflect complexity (Sorenson et al., 2006).

The index of *Balland and Rigby* rests on the assumptions that complex knowledge is relatively scarce geographically and that it tends to co-concentrate with other complex knowledge in space (Balland and Rigby, 2017). However, the spatial distribution of knowledge may have multiple explanations, including complexity. For instance, the diffusion of knowledge in space and, hence, its geographic distribution, depend on its degree of maturity, popularity, natural conditions, geographic distance, place of origin, and (again), crucially, economic potential (Hägerstrand, 1967; Teece, 1977; Rogers, 1995; Zander and Kogut, 1995).[3]

Lacking an objective criterion of knowledge complexity, it is difficult to assess the severity of these assumptions. Hence, the extent to which these measures actually capture what they are intended to capture is unclear.

The present paper contributes to the literature with a novel approach to this matter. Based on a conception of technologies as combinations of (knowledge) components (Fleming and Sorenson, 2001; Hargadon, 2003), it uses insights from complexity and network science to introduce *structural diversity* as a new measure of technological complexity. It represents the diversity of (subnetwork) topologies in technologies' combinatorial networks. The paper also puts forward that the Network Diversity Score developed by Emmert-Streib and Dehmer (2012) to quantify the complexity of networks, can serve as an empirical approximation of this measure.

Using this approach, the complexity of 655 technologies is quantified on the basis of EU patent data from 1980 to 2015. The results are shown to correspond to a number of stylized facts, which the literature suggests characterize technological complexity: Complexity is increasing over time, it requires more R&D and it involves more collaboration. In contrast, the stylized fact of complex technologies to concentrate in space is only confirmed when controlling for the size of technologies.

[2]To the best of the author's knowledge, *Fleming and Sorenson*'s measure has not been used to evaluate complexity at the level of technologies. However, in principle, its patent-specific complexity values can easily be aggregated to the level of technologies.

[3]From an empirical perspective, constructing a complexity index on the basis of the spatial distribution of knowledge raises two additional issues: It represents a potentially endogenous variable in many spatial research settings and its values are conditional on the delineation of the employed spatial units.

1.2 Technologies as combinatorial networks

In management science and engineering, *technologies* are described as systems of interrelated components (Hargadon, 2003; Arthur, 2009; Mcnerney et al., 2011). Components are all "*parts of a technology or the steps in an industrial process*" (Mcnerney et al., 2011, p. 9009), and two components are related if changes in one of them affects the respective other. A comparable conceptualization can be found in innovation studies. In this literature, *technologies* are described as sets of interrelated components with the latter being knowledge "pieces" and relations being their combinations (Sorenson and Fleming, 2004). For example, "*one might think of the automobile as a combination of the bicycle, the horse carriage, and the internal combustion engine*" (Fleming and Sorenson, 2001, p. 1020). While the conceptualization in engineering is instead focused on technological systems, with the set of relations between components being known as *design structure matrix* (Steward, 1981), innovation studies focus on the knowledge dimension of technologies. In this literature, a technology is described as a *recipe*, which encompasses information on constituent knowledge components and their combinations (Sorenson and Fleming, 2004). Crucially, both views apply a network perspective with nodes representing components and links representing their relations/combinations. In these networks, some components are directly linked while others are indirectly related. For instance, in regard to airplane technology, the components' wing design and aluminum processing are directly linked, while electronic navigation is only indirectly related since other components (e.g., electronic control systems, mechatronic interfaces) act as bridges.

The network perspective allows the assessment of technologies' complexity based on the combination of their components (Fleming and Sorenson, 2001; Sorenson and Fleming, 2004; Mcnerney et al., 2011). The measure of *structural diversity* extends this approach with insights from network science. In such research, a wide range of measures has been put forward that quantify the structural complexity of networks (Bonchev and Buck, 2005; Dehmer and Mowshowitz, 2011). Importantly, some measures evaluate networks' complexity from the perspective of information theory (Wiener, 1947; Shannon, 1948). In essence, information theoretical measures of network complexity quantify the amount of information contained in networks, i.e. the quantity of information needed to describe the full network, or alternatively to describe its most important structural features (e.g., their degree distribution). In any case, a network's complexity is argued to increase with growing amounts of information needed for its description (Dehmer et al., 2009).

Lets assume a technology's combinatorial network has a star-like structure (Figure 1 a). This network, and thereby the technology, can be easily described once this structure and the central knowledge component are known. In this case, the technology only consists of this central knowledge component and its direct combination with all other components. From an information theoretical perspective, the description of this technology requires minimal information because there is only one (simple) network structure dominating its combinatorial network. The low information content makes it likely that this technology can be easily invented, copied, and codified. An (oversimplified) example of such a technology with a star-like combinatorial network is a table. It usually consists of five components (four poles and one table board). Each pole is directly and exclusively connected to the table board, which therefore constitutes the central component. From the information theoretical perspective, the combinatorial network of the table indicates it is a simple technology.

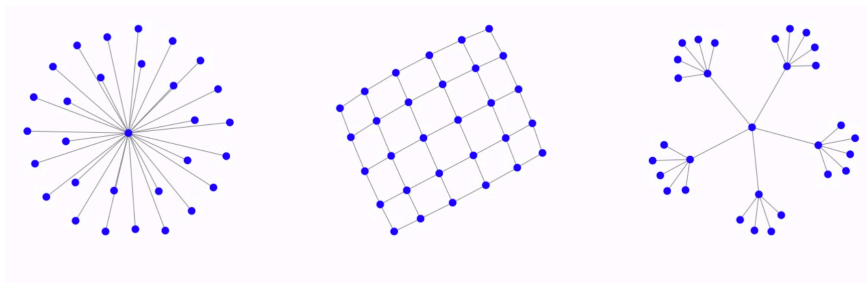
Similarly, only a limited amount of information is required to describe a lattice network (Figure 1 b), as its degree distribution already contains most of its structural characteristics. More information is contained in a tree-like network (Figure 1 c). Here, in order to be described, the identity of multiple central nodes and the depth of its hierarchy are needed. The amount of information to represent a small-world network (Figure 1 d) is even larger. In fact, small-world structures are usually used as examples for complex networks (Emmert-Streib and Dehmer, 2012).

What differentiates simple networks from more complex ones is the existence of certain kinds of organizational principles underlying their structures. These principles allow for condensing the information required to describe a network.[4] The existence of organizational principles usually translates into specific (sub)network structures (e.g., stars,

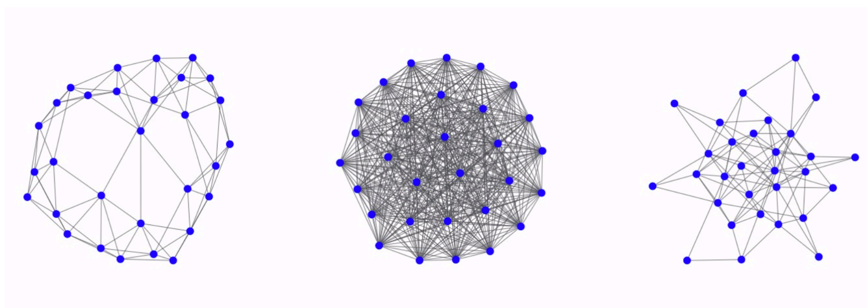
[4]Frequently, these structuring principles are the result of specific network formation mechanisms, such as preferential attachment or transitivity.

lattices, cliques, hierarchies) that appear in a network. These are called network *topologies* in the remainder of the paper, and they are the basis for the proposed approach for measuring technological complexity.

In line with an information theoretical view on network complexity, I argue that the more information is required to describe the topology of a technology's combinatorial network, the more complex it is. Moreover, multiple topologies will usually be needed to describe a combinatorial network's structure, as these networks may consists of subnetworks, each with different topologies. In other words, it is unlikely that a technology's components will connect in just one way (e.g., in a star-like manner). Alternatively, topologies may also be mingled. For instance, a small-world network combines tree-like and clique-like topologies. Since each topology's description implies additional information, the total information required to describe the full network will increase the more distinct topologies are present. With increasing information, the complexity of the combinatorial network grows along with the complexity of the corresponding technology. Consequently, a quantification of this diversity in topologies can be seen as a measure of technological complexity, which will be called *structural diversity* in the following. Notably, according to this view, a random network (Figure 1 f) is the most complex theoretically because it consists of the largest number of distinct topologies.



(a) Star (b) Lattice (c) Tree



(d) Small-World (e) Complete (f) Random

Figure 1: Archetypal network topologies

2 Method

2.1 The measure of structural diversity

There is no commonly accepted method of assessing the complexity of networks or the diversity in its sub-structures and topologies. It is beyond the scope of the present paper to review or discuss the pros and cons of the many existing approaches, as this can be found elsewhere (Bonchev and Buck, 2005; Dehmer and Mowshowitz, 2011; Emmert-Streib and Dehmer, 2012).

Recently, the *Network Diversity Score* (NDS) was introduced and compared to other common measures of network complexity (Emmert-Streib and Dehmer, 2012). In contrast to these measures, only the NDS is capable of consistently separating ordered, complex, and random networks. Networks are considered *ordered* when many nodes exhibit similar properties (e.g., degree), which corresponds to one or a few dominant topologies. According to the previous discussion, ordered combinatorial networks represent relatively simple technologies because of their comparatively more homogeneous topologies. *Complex* networks represent mixtures of *ordered* and *random* structures. They are therefore characterized by a larger topological heterogeneity compared to *ordered* networks. For example, a small-world network is usually seen as complex (Emmert-Streib and Dehmer, 2012) because it involves multiple topologies such as stars and triangles of different sizes (Figure 1 d). Random networks are the most structurally diverse as they involve the largest heterogeneity of topologies. Despite the variance in specific network topologies used in its definition, the NDS does not directly measure the structural diversity of networks. However, it ranks networks on a scale ranging from ordered over complex and to random networks, which empirically represents the idea of *structural diversity*.[5]

The NDS differs in multiple ways from traditional measures of network complexity. Firstly, it is a result of scientific numerical experimentation. More precisely, while based on some general theoretical ideas on what characterizes simple and complex networks, the measure is empirically optimized to significantly differentiate between artificially created random, ordered, and complex networks (Emmert-Streib and Dehmer, 2012). Secondly, the measure combines multiple network characteristics into one: It considers the share of modules ($\alpha_{module} = \frac{M}{n}$) with M being the number of modules and n being the number of nodes. Modules are densely connected subgraphs in a network. The variance of module sizes m $v_{module} = \frac{var(m)}{mean(m)}$ is also included. Random networks are likely to show a low variability and low average size of modules. Further, the variable V_λ captures the Laplacian (L) matrix's variability defined as $v_\lambda = \frac{var(\Lambda(L))}{mean(\Lambda(L))}$. Lastly, the relation of motifs of sizes three and four enters the measure. This variable is observed to be highest in ordered networks, medium in complex networks, and lowest in random networks (Emmert-Streib and Dehmer, 2012). Counting the number of motifs in networks usually implies concentrating on those network three- and four-node structures that are overrepresented in the empirical network in comparison to a random network (Milo et al., 2002). Due to the substantial computational burden of the randomization of all sample networks, I adapt this part of the NDS measure and replace the motifs-based relation with the ratio of graphlets of sizes three and four. Hence, I estimate the ratio $r_{graphlet}$ between all empirically observed network structures based on three nodes (graphlets of size three, $N_{graphlet}(3)$) and those involving four nodes (graphlets of size four, $N_{graphlet}(4)$) as $r_{graphlet} = \frac{N_{graphlet}(3)}{N_{graphlet}(4)}$.

The four variables are combined in the individual network diversity score ($iNDS$) of the network (G_T):

$$iNDS(G_T) = \frac{\alpha_{module} * r_{graphlet}}{v_{module} * v_\lambda}. \tag{1}$$

Networks may show properties of a complex or ordered network merely by chance and thereby mislead measures of complexity. Therefore, $iNDS$ is estimated for a population of networks G_M to which G_T belongs (Emmert-Streib and Dehmer, 2012). In practice, drawing random samples S from network G_T and estimating $iNDS$ for each sample network achieves this. The final network diversity measure (NDS) is obtained by:

$$NDS(\{G_T^S|G_M\}) = \frac{1}{S}\sum_{G_T \epsilon G_M}^{S} iNDS(G_T) \tag{2}$$

To allow for an easier interpretation, I transform the measure such that large values signal random networks (complex technologies), medium values indicate complex networks (medium complex technologies), and low values represent ordered networks (simple technologies). This is done by taking NDS in logs and subsequently multiplying it by -1. The obtained value represents the structural diversity of a technology's combinatorial network and will be denoted as *structural diversity* in the remainder of the paper. Notably, its values may vary somewhat when estimating it repeatedly for the same technology due to the random sample selection procedure.

[5]However, this does not suggest that technologies with random combinatorial networks actually exist or are even possible. Ultimately, the cumulative character of technological development and the relevance of social processes underlying it will ensure the presence of systematic structures.

2.2 Data

Using the measure of *structural diversity* I estimate the complexity of technologies on the basis of patent data. Despite well-known problems (Griliches, 1990), patents entail detailed and unparalleled information about technologies and their innovation processes. I use the OECD REGPAT database (version 2018) covering patent applications to the European Patent Office. As there is a time lag between the priority date and the availability of patent information, the most recent years of this data are unreliable. The analysis is therefore restricted to the years 1980 to 2015. It utilizes information on $3,137,881$ patent applications. They are assigned to countries and regions by means of inventors' residences (multiple-counting). Technologies are defined on the basis of the *Corporate Patent Classification* (CPC). The CPC is hierarchically organized into nine classes at the highest level and into more than 230,300 subclasses at the lowest level. I use the four-digit CPC level to define 655 distinct technologies. While there is no objective reason for this level, it offers a good trade-off between technological disaggregation and manageable numbers of technologies. In addition, it has been used in related studies (Schmoch et al., 2003; Breschi and Lenzi, 2011).

Patent numbers vary considerably between years and some technologies have few patents. Therefore, a moving window approach is used to calculate annual complexity measures. In other words, I combine the patent information of three years such that a technology's complexity measure in year t is based on patents issued between years t and $t-2$.[6]

2.3 Calculating structural diversity

When applying the structural diversity measure to patent data, technologies' (knowledge) components and their combinations must be defined. I consider the lowest level of CPC classes (10-digit subclasses) as approximations of components and their co-occurrence on patents as combinations (Fleming and Sorenson, 2001; Sorenson and Fleming, 2004). The estimation is done on this basis as follows for each year t (moving window of three years): First, for each of the 655 technologies T, all patents are extracted with at least one of their 10-digits CPC subclasses belonging to the four-digit class of the focal technology ($Pats_T$). Second, the matrix M_T is created from all co-occurrence counts of the (ten-digit) CPC subclasses assigned to the patents $Pats_T$. M_T is dichotomized with all positive entries set to one. The dichotomization is necessary because the NDS measure is not (yet) defined for valued networks.[7] M_t is an adjacency matrix representing the network G_T of all of the ways technology T's components have been combined among themselves, how they have been combined with other technologies' components, and how other technologies' components are combined when at least one component of technology T is involved. Hence, it summarizes all (knowledge) combinations related to technology T, i.e. a technology's combinatorial network.[8]

The *structural diversity* of technology T is obtained by applying the NDS measure to network G_T. However, as the NDS requires connected networks (Emmert-Streib and Dehmer, 2012), the estimation is restricted to the main component of network G_T. While in early years (<1985), the largest component represents less than 50% of the combinatorial networks' nodes, its size rises quickly and on average represents more than 75% of the nodes by 1997 (see Figure 10 in the Appendix). For each G_T (main component), a set of $S = 50$ nodes is randomly selected. For each node c ($c \in S$), a network $G_{T,c}$ is drawn from G_T by a random walktrap of $n = 150$ steps starting from c.[9] The $iNDS$ (Equation 1) is then calculated for all sample networks $G_{T,c}$ and subsequently averaged (Equation 2). The results is denoted as NDS_T and it represents the empirical measure of *structural diversity* of the combinatorial network

[6] Choosing moving windows of three years is rather arbitrary but represents a good trade-off between the smoothness of the development and the temporal variance of the measure.

[7] This simple dichotomization also keeps most of the original (valued) network's information. Nevertheless, future work should explore the effect of alternative dichotomization approaches on the measure.

[8] Alternatively, the network can be restricted to component combinations exclusively involving technology T. However, such a restriction would ignore potential bridging functions of adjacent technologies and how technology T is embedded into the overarching technological space.

[9] In case the network has fewer than 50 nodes, S was set to its number of nodes. The choice of S and n represent a trade-off between robustness and computational burden. Previous research found a sample size of $S = 10$ subnetworks with a size of $n = 120$ nodes to be sufficiently robust for comparable real-world networks (Emmert-Streib and Dehmer, 2012).

of technology T. The measure is separately calculated for each technology in every year t resulting in a year and technology specific complexity value. That is, the multi-step procedure of calculating NDS_T is repeated 655×38 (technologies $\times$ years) times.

3 Empirical analysis

The presentation of the empirical results is centered on four stylized facts of technological complexity that most scholars in the field seem to agree upon: Technological complexity increases over time, complex technologies involve more R&D and require more collaboration. Moreover, complex technologies tend to concentrate in space.

3.1 Technological complexity increases over time

Figure 2 presents the distribution of *structural diversity* of the 655 technologies in each year between 1980 and 2015. The observed minimum is zero and the maximum is 14.98. In between, the distribution is bimodal with a peak at zero and a maximum density at moderate values. The peak at zero reflects many technologies having no or too few patents to calculate *structural diversity*.[10] Due to generally rising patent numbers, this peak becomes less pronounced over the years. When abstracting from the peak at zero, the distribution is bell-shaped with short right-hand and somewhat longer as well as over time growing left-hand tails. Otherwise, the general shape remains relatively similar over time. The relative stability of technologies' rankings is confirmed by the measure's temporal (rank) correlation (Figure 9 in the appendix). With few exceptions, the median increases over the years, which gives a first impression

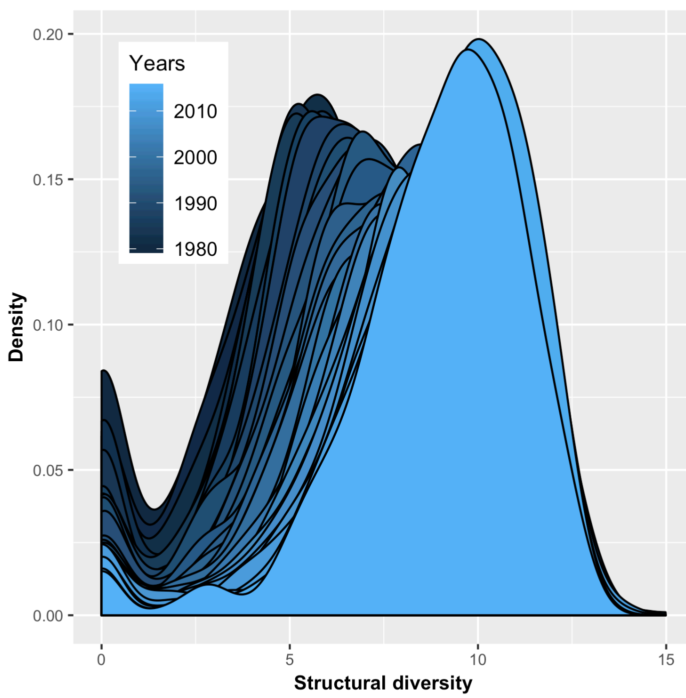


Figure 2: Distribution of structural diversity 1980-2015

of *structural diversities'* temporal development. Technological complexity is argued to increase over time due to knowledge and technologies' cumulative natures, thus implying that each generation is building upon the technological environment established by its predecessors (Nelson and Winter, 1982; Howitt, 1999; Aunger, 2010; Hidalgo, 2015). Technologies also become more complex due to their growing range of functions. For instance, "[d]*igital control*

[10]The main component of the combinatorial needs be at least of size two.

systems [of aircraft engines] *interact with and govern a larger (and increasing) number of engine components than* [previous] *hydromechanical ones*" (Prencipe, 2000, p. 904). Another example is Microsoft's operation system Windows, which grew from 3-4 million lines of code (Windows 3.1) to more than 40 million (Windows Vista) (Wikipedia, 2017). Moreover, technologies have reached higher levels of complementarity requiring more multi-technology activities, which adds to the complexity of their development and application (Fai and Von Tunzelmann, 2001). In summary, "[t]*he result is a constantly increasing sophistication and richness of the technological world*" (Aunger, 2010, p. 773).

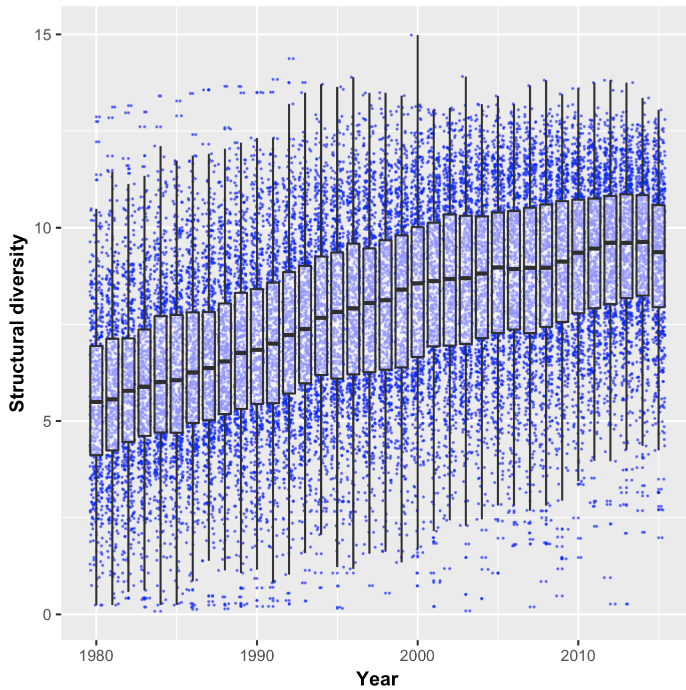


Figure 3: Development of structural diversity over time (1980-2015)

The boxplots in Figure 3 reveal the median *structural diversity* of the 655 technologies to grow over time, which is in line with the argument of increasing technological complexity.[11] Notably, the variance of *structural diversity* remains high with the lowest values observed for 2015 being well below the median in 1980. Moreover, a set of technologies already reaches values in the early 1980s larger than the highest values in most recent years. However, very low patent numbers characterize these technologies, which makes the patent-based assessment of their technological complexity less reliable.

Another way to examine the evolution of complexity over time is to compare young and old technologies. For this comparison, the age of all patents has been calculated by subtracting their priority year from the most recent year in the data (2015). A technology's age in year t is then represented by the median of the age of all patents (at least one of their CPC subclass belongs to this technology) that have been granted in t or before. The rank correlation coefficient of technologies' *structural diversity* value and their patents' median age are plotted in Figure 4 for each year. With the exception of 1980 to 1983, the correlation is significantly negative, thus signaling that younger technologies obtain higher values of *structural diversity*. From 1992 onwards, this correlation is very strong, with the coefficient fluctuating around $r = -0.47$. The finding implies that the increase of technological complexity over time is partly explained by new technologies being more complex than older ones.

[11]Note that a jitter algorithm has been used to distribute the dots for maximizing visibility.

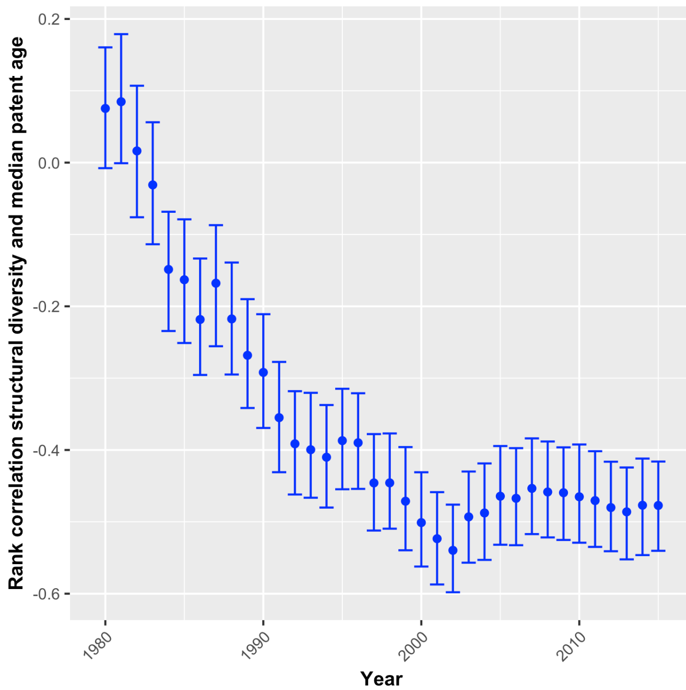


Figure 4: Correlation between structural diversity and technologies' age, 1980-2015

3.2 Complex technologies require larger R&D efforts

Technologies are advanced by creating new knowledge combinations through search activities for potentially fitting pieces and subsequent testing of these combinations, which is frequently done by trial-and-error (Carbonell and Rodriguez, 2006). "Harder-to-find," i.e. more difficult/complex solutions, involve more trials and errors, which consume resources. Complex technologies are based on greater knowledge diversity and on the combination of less common knowledge than simple technologies (Fleming and Sorenson, 2001), which further increases the efforts needed in development processes. Additionally, learning complex knowledge is more resource-intensive because greater absorptive capacities are needed (Cohen and Levinthal, 1990) and passive learning modes are insufficient (Pintea and Thompson, 2007). These features of complex technologies translate into longer development times for complex products (Griffin, 1997). In line with this, organizations are more likely to fail when engaged in the development of complex technologies (Singh, 1997). At the national level, R&D intensity is moreover observed to outgrow economic outputs and incomes because of increasing complexity and development diversity (Pintea and Thompson, 2007; Kim, 2015). In sum, the development of complex technologies requires more R&D efforts than simpler technologies.

Unfortunately, there is hardly any information on R&D efforts available that can be matched to the employed patent data. I therefore use two alternative approximations, none of which is perfect: patents and being classified as high-tech. Patents and R&D efforts are positively correlated at the organizational and regional levels (Griliches, 1990; Acs et al., 2002) suggesting that total R&D efforts are larger in technologies with many patent applications.[12]

Figure 5 shows the rank correlation coefficients for the years 1980 to 2015 of *structural diversity* and the number of patents assigned to a technology. The correlation coefficient is strongly positive and significant in all time periods ranging from $r = 0.45$ (1980) to $r = 0.69$ (2015). To the extent that patent numbers reflect R&D intensity, it confirms the positive relation between complexity and R&D intensity. The positive trend in the coefficient's development suggests that this relationship intensifies over time - i.e. reaching higher levels of complexity is more dependent on

[12]Note, however, that there are also considerable differences in industries' patent propensities (Arundel and Kabla, 1998).

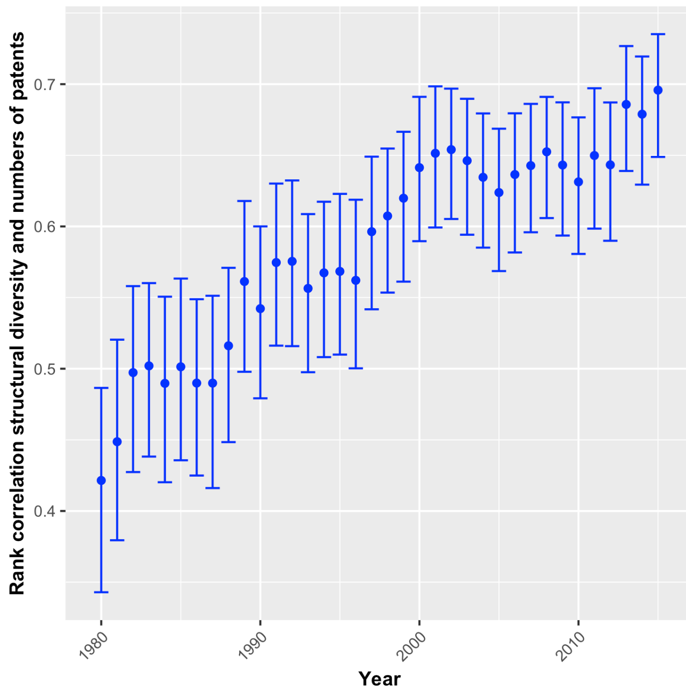


Figure 5: Correlation between structural diversity and technologies' patent counts, 1980-2015

R&D investment than in the past. Potentially, this trend reflects the diminishing returns to R&D hypothesis according to which innovations are increasingly distributed across more products, which results in declining returns to R&D over time (Madsen, 2007).

As an alternative measure of R&D efforts, I compare high-tech and non high-tech technologies. High-tech research is characterized (and frequently defined) by larger R&D efforts and intensity (Mendonça, 2009). It is also directly linked to complex technologies (Coad and Rao, 2008; Sáenz et al., 2009). Accordingly, technologies considered high-tech are expected to obtain larger values of *structural diversity* than other technologies.

The Trilateral Statistical Report from the European, Japanese, and US patent offices identifies 31 four-digit patent subclasses as high-tech (EPO et al., 2007). High technologies include the fields of computer and automated business equipment, aviation, microorganism and genetic engineering, lasers, semiconductors, and communication.

Figure 6 compares the *structural diversity* of high technologies with all other technologies. With the exception of one year (1981), high technologies show on average larger values of *structural diversity*. The difference becomes statistically significant in the late 1980s. Hence, the measure of *structural diversity* also identifies high technologies as more complex, which provides further support for the argument that complex technologies require larger R&D efforts. Nevertheless, these results should not be over-interpreted because patent counts and belonging to high technology are far from being precise approximations of R&D intensity.

3.3 Complex technologies require more collaboration

Another feature commonly associated with complex technologies is their greater need for collaboration in R&D (Hidalgo and Hausmann, 2009; Hidalgo, 2015; Balland and Rigby, 2017). In particular, the larger knowledge diversity inherent to complex technologies demands more diverse but specialized experts (Pavitt, 1998); experts who must work together to solve complex problems (Carbonell and Rodriguez, 2006).

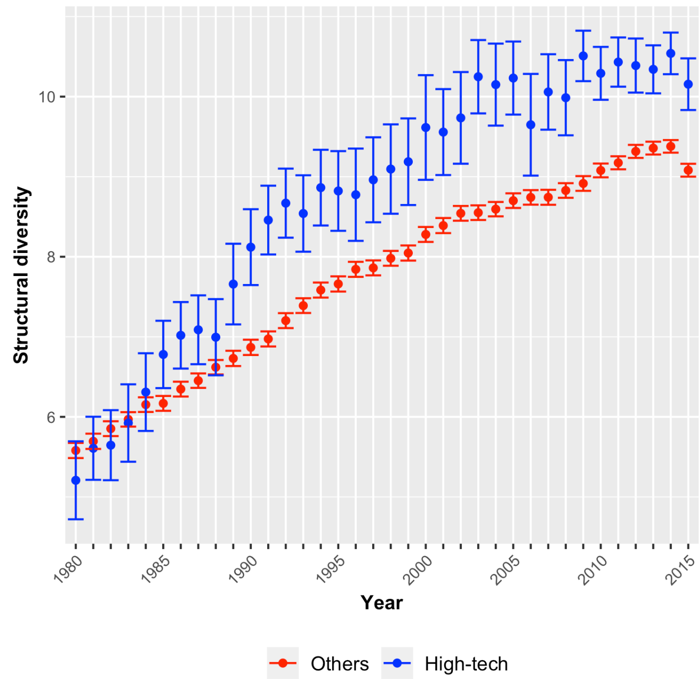


Figure 6: Structural complexity of *high-technologies*, 1980-2015

Figure 7 shows the rank correlation between technologies' *structural diversity* and the average number of inventors per patents. The latter signals the extent to which the patent is based on teamwork. The positive significant coefficient in all years underlines the importance of collaborative work in more complex technologies. While the correlation is relatively low in the beginning of the 1980s, it grows strongly, reaching a value of approximately 0.4 in the late 1980s. It remains somewhat above or close to this level in subsequent years. The finding clearly supports complex technologies involving more collaboration in R&D.

3.4 Complex technologies concentrate in space

In economic geography and regional science, it has long been argued that developing complex technologies requires special skills, existing expertise, infrastructure, and institutions not found everywhere (Jaffe, 1989; Audretsch and Feldman, 1996; Almeida, 1996). Spatial proximity between experts is essential for face-to-face communication, which enhances work on complex projects (Carbonell and Rodriguez, 2006). The place-specificity of favorable conditions for (complex) innovation is also emphasized in concepts like the *learning regions*, *innovative milieu*, and *regional innovation systems* (Florida, 1995; Camagni, 1991; Cooke, 1992). Such conditions allow for bridging cognitive distances and combining heterogeneous knowledge, which in other places would remain uncombined. These place-specificities are path-dependent and relatively rare. Consequently, complex technologies are argued to concentrate in space, which is supported by empirical evidence for the USA (Balland and Rigby, 2017; Sorenson, 2005).

To explore the relation between *structural diversity* and the spatial distribution of technologies, the residential information of patent inventors is used. More precise, for each NUTS2 region and technology, I count the number of patents with at least one of its inventors' addresses being assigned to this region. Subsequently, Gini coefficients are estimated for the technology-specific regional distributions of patent counts (270 regions and $1,557,416$ patents).[13] The coefficient obtains values close to one if inventors concentrate in a few regions and it converges to zero when they

[13] The spatial concentration of technologies' regional patent counts are calculated with respect to European NUTS2 regions. Consequently, only patents are considered with at least one inventor from Europe.

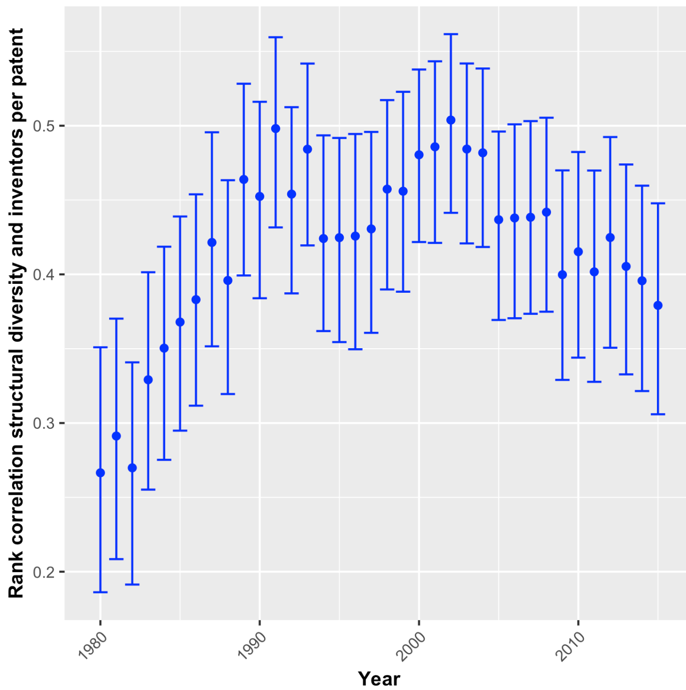


Figure 7: Correlation between *structural diversity* and the number of inventors per patent, 1980-2015

are evenly distributed in space. Figure 8 shows the correlation between technologies' spatial Gini coefficients and their values of *structural diversity*. As technologies with few patents have less potential to be equally distributed across regions, the correlation is also presented for technologies with at least $1,000$ patents.

When considering all technologies (blue error bars), the correlations are strongly negative significant and suggest that complex technologies are more evenly distributed than simple technologies. However, the picture changes when concentrating on technologies with many patents (red error bars). Due to the small numbers of technologies with more than 1,000 patents, the correlations remain insignificant until the mid-1990s. From that year onward, the correlations fluctuate between being positive significant and insignificant. Notably, since 2009, the coefficient grows constantly and remains significant from 2012 onward, suggesting that larger and more complex technologies do increasingly concentrate in space. However, a similar trend was visible in the 1990s, which turned around in 1999. It is therefore not clear if it indicates systematic changes in the underlying processes or merely empirical fluctuations. Hence, the evidence is rather inconclusive with only the largest and most patent intensive technologies showing some spatial concentration in the most recent years.

3.5 Multivariate analysis

Thus far, the empirical analysis has analyzed differences between simple and complex technologies in a bivariate manner. This does not deliver insights into the relative importance of some of these differences or into the extent that they are related. Table 1 presents the results of a linear (within-estimator with time-fixed effects) panel regression with the *structural diversity* values of 646 technologies (those with at least five patents in two subsequent years) over the 36 years (1980-2015) as the dependent variable. All none-dummy explanatory variables are considered in logs, and robust clustered standard errors are used. As a robustness check, I repeat the analysis limiting the sample to the most recent 15 years. These results are reported in Table 5 in the Appendix. No substantial differences are observed between the two analyses. I therefore concentrate on the findings of the analysis based on all years in the following. The Appendix also includes an overview of the variables' descriptives (Table 3) and their correlations (Table 4).

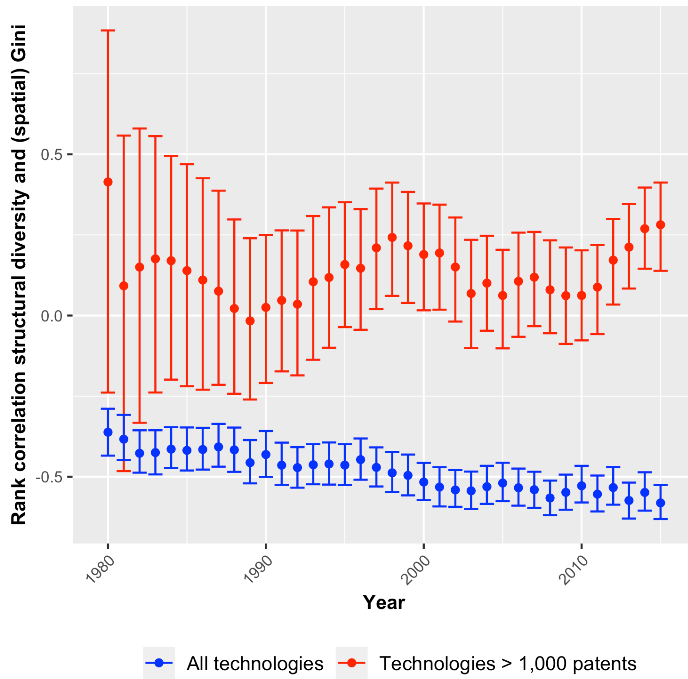


Figure 8: Correlation between *structural diversity* and technologies' spatial concentration, 1980-2015

	Dependent variable:							
	Structural complexity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Patents)	1.003***	1.002***	0.913***	0.827***	0.897***		1.177***	0.903***
	(0.036)	(0.037)	(0.039)	(0.037)	(0.053)		(0.051)	(0.038)
High		0.030	−0.148	−0.346*	−0.390**	0.842***		−0.132
		(0.252)	(0.228)	(0.197)	(0.197)	(0.325)		(0.148)
Log(Median age + 1)			−2.324***	−2.032***	−1.977***			−1.083***
			(0.266)	(0.239)	(0.239)			(0.182)
Log(Inventors per patent)				2.888***	2.837***			0.980***
				(0.198)	(0.197)			(0.137)
Log(Spatial Gini)					2.000**	−14.828***	5.195***	4.196***
					(0.988)	(1.039)	(1.102)	(0.630)
log(CPCs per patent)								2.380***
								(0.069)
adj. R2	0.396	0.396	0.421	0.485	0.486	0.177	0.404	0.653
n	646	646	646	646	646	646	646	646
T	36	36	36	36	36	36	36	36
N	22,360	22,360	22,360	22,360	22,344	22,344	22,344	22,310
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Unbalanced panel regression, robust standard errors and p-values. *p<0.1; **p<0.05; ***p<0.01

Table 1: Characteristics of *structural diversity*, 1980-2015

The multivariate analyses confirm the previous bivariate results. More complex technologies tend to have more patents, which is underlined by the significantly positive coefficient of *Patents*. Hence, they are likely to require larger R&D efforts. More complex technologies are also younger, as the coefficient of *Median age* is significantly negative. The significantly positive coefficient of *Inventors per patent* confirms that R&D in complex technologies is conducted in a more collaborative fashion than in simple technologies. Accordingly, these characteristics of technological complexity are found to be true even when controlling for the respective others. Thus, to a significant extent, they characterize complex technologies independent of each other. This cannot be said about technologies being high-tech and their spatial distribution. The dummy for high technologies (*High-tech*) is strongly correlated to other explanatory variables, particularly *Median age* and *Patents*, which explains its insignificance in most models. The weakly significant negative coefficients of *High-tech* in Models (4) and (5) indicate that high technologies have relatively lower values of *structural diversity* when controlling for the average age of their patents, collaboration intensity, and patent numbers. When excluding these variables, the coefficient shows the expected significant positive sign. Accordingly, the larger complexity of high technologies seems to be primarily explained by these other features.

The measure of spatial concentration (*Spatial Gini*) also relates very strongly to the other explanatory variables in general and (negatively) to the number of patents in particular (see Table 4 in the Appendix). When controlling for the number of patents, it becomes significantly positive (Model 7). However, its significance is somewhat reduced when including the other explanatory variables (Model 5). Hence, patent numbers largely explain the observed negative bivariate relationship between *structural diversity* and technologies' spatial concentration. More patents tend to make technologies more evenly distributed in space. Once this is accounted for, technologies with large values of *structural diversity* (complex technologies) are found to concentrate in space.

Lastly, to control for potential changes in the classification of patents to CPC classes, the number of CPC subclasses (10-digit) per patent (*CPCs per patent*) is added to the model. The variable becomes positive and significant in the model and primarily seems to lower the explanatory power of patents and technologies' age. While the number of subclasses assigned to patents tends to positively correlated to *structural diversity*, all other results do not change substantially.

3.6 Comparison with two alternative measures of technological complexity

To put these results for the measure of *structural diversity* into perspective, I repeat the multivariate analysis for the two alternative approaches of quantifying technological complexity that have been or can be used in similar settings. The first is the complexity measure of *modular complexity* ($FS.modular$) introduced by Fleming and Sorenson (2001). In the present paper, it evaluates the frequency of patent subclass co-occurrences (10-digit CPC classes) on patents in a particular year (moving window of three years), in comparison with the cumulative frequency of their co-occurrences in all prior (to the moving window) years. The individual scores of patents are averaged (median) at the four-digit CPC level. Secondly, I follow Balland and Rigby (2017) in calculating an index of technological complexity (KCI) based on technologies' spatial distribution in year t. For this, the regional technological advantage (RTA) is calculated for all European regions (NUTS 2) and technologies (four-digit CPC).[14] On this basis, a two-mode network between regions and technologies T is constructed with a binary link if region r has $RTA_{r,T,t} > 1$, i.e., when it is above average specialized in technology T. There is no link otherwise. The method of reflection with 20 iterations is applied to this network generating the complexity index KCI for year t. To resemble the construction of the *structural diversity* measure, the three-year moving window approach is employed in the construction of the annual patent data. The KCI is an index by construction, which requires the use of year-fixed effects in the regression to make it comparable across years. Moreover, it is advisable to transform it into ranks to control for annual variations in its variance. I estimate a regression for both versions, whereby log-transforming the ranks-based KCI substantially improves the model fit. However, the regression results obtained for the original and the ranks-based version of the KCI are identical in terms of the coefficients' signs and significance.

The analysis (Table 2) reveals negative relations with R&D efforts (approximated by patents and belonging to high-tech) and collaborative R&D when approximating technologies' complexity with the KCI. Older (*Median age*) and less collaborative technologies (*Inventors per patent*) are also found to score higher on this complexity index. A significantly

[14] The results did not change substantially when using NUTS 3 regions.

positive relation is observed with spatial concentration (*Spatial Gini*). With exception of the latter, the results suggest this index to behave rather opposite to the stylized facts usually associated to technological complexity. In case of $FS.Modular$, the findings are somewhat more in line with these facts. The measure is found to be positively related to R&D efforts (*Patents*, *High-tech*), to collaborative R&D (*Inventors per patent*), and to the spatial concentration of patenting activities (*Spatial Gini*). However, similar to KCI, older technologies (*Median age*) are associated to higher levels of complexity, which contrasts the corresponding stylized fact.

Table 2: Characteristics of alternative measures of technological complexity, 1980-2015

	Dependent variable:			
	Complexity			
	FS.7838787Modular	HH.NUTS2	log(HH.NUTS2.rank)	HH.NUTS3
	(1)	(2)	(3)	(4)
log(Patents)	0.077***	-0.225^{*}	-0.074^{***}	-0.367^{***}
	(0.015)	(0.118)	(0.021)	(0.095)
High-tech	0.115*	-1.810^{***}	-0.244^{***}	-1.153^{***}
	(0.063)	(0.484)	(0.086)	(0.264)
Log(Median age)	0.262***	0.630	-0.115	0.406
	(0.065)	(0.581)	(0.096)	(0.440)
Log(Inventors per patent)	0.107	-3.810^{***}	-0.595^{***}	-5.768^{***}
	(0.068)	(0.428)	(0.078)	(0.342)
Log(Spatial Gini)	2.553***	26.493***	4.711***	16.651***
	(0.305)	(2.127)	(0.392)	(1.753)
adj. R2	0.062	0.125	0.208	0.154
n	646	646	646	646
T	36	36	36	36
N	21,974	22,355	22,355	22,355
Year fixed effects	Yes	Yes	Yes	Yes

Unbalanced panel regression, robust standard errors and p-values. *p<0.1; **p<0.05; ***p<0.01

This exercise is not intended and surely does not qualify as a fully developed comparison of the different approaches. However, it highlights the non-arbitrary character of choosing an indicator of technological complexity. Context matters, and, in some situations, certain features of an approach are desirable while in others they might be misleading. If it is important to have a measure reflecting the four stylized facts discussed above; of these three, *structural diversity* seems to mirror them most closely.

4 Summary & Conclusion

Measuring the complexity of technologies has received significant attention from different disciplines. For instance, in engineering, technological complexity is argued to impact the costs and the management of technological systems (Mcnerney et al., 2011). Scholars use concepts and measures of technological complexity to better understand combinatorial R&D processes in innovation studies (Fleming and Sorenson, 2001; Sorenson and Fleming, 2004). Moreover, in economics and economic geography, technological complexity is seen as an important determinant of the uneven economic development in space (Hidalgo and Hausmann, 2009; Balland and Rigby, 2017). However, quantifying the complexity of technologies is literally a *complex* task and there is no widely accepted way to do it.

The present paper proposed a new measure of technological complexity called *structural diversity*, which approximates the diversity in how technologies' (knowledge) subcomponents relate to each other. It was also argued that a slightly adapted version of the *Network Diversity Score* by Emmert-Streib and Dehmer (2012) resembles this measure in empirical settings. Employing this approach, the study assessed the complexity of 655 patent classes (technologies) across 36 years. Co-occurrences of CPC subclasses on patents have been used to create technology-specific combinatorial networks, which, in turn, served as basis for the calculation of *structural diversity*. Table 6 in the Appendix provides the complete list of these technologies and their respective complexity values in the year 2014.[15]

Subsequently, it was shown that the obtained values mirror four stylized facts commonly associated with technological complexity: Complexity growth over time with complex technologies being on average younger. These technologies are also more R&D intensive and their R&D activities are more collaborative. When accounting for technologies with many patents being more widely distributed in space, complex technologies have moreover been shown to concentrate geographically.

The present study focused on the introduction of a new measure of technological complexity and explored its properties empirically. While the obtained results are promising, the empirical measurement of *structural diversity* calls for more work in the future. The application of the *Network Diversity Score* is an approximate measure of the diversity of network topologies. In addition, it is relatively computational intensive and requires the dichotomization of technologies' combinatorial networks. The latter aspect particularly gives direction for future research, as it implies a significant loss of information contain in the patent data.

Further, the presented empirical analysis is only a first step towards a better understanding of the development of technological complexity over time and space as well as of its relation to socio-economic developments. For instance, when using this measure in future studies, it will be interesting to investigate the relevance of technological complexity for the economic growth of firms, regions, and countries. Similarly, evaluating the contribution of policy and public research to the advancement of simple and complex technologies will offer new insights into their over-all role in technological progress.

Acknowledgments

The author would like to thank Lars Mewes, Louis Knüpling, Pierre-Alexandre Balland, and the anonymous reviewers for their helpful comments and suggestions.

References

Acs, Z. J., Anselin, L., and Varga, A. (2002). Patents and Innovation Counts as Measures of Regional Production of New Knowledge. *Research Policy*, 31:1069–1085.

Almeida, P. (1996). Knowledge sourcing by foreign multinationals : Patent citation analysis in the U.S. semiconductor industry. *Strategic Management Journal*, 17(S2):155–165.

Arthur, B. (2009). *The Nature of Technology: What It Is and How It Evolves*. Free Press, London.

Arundel, A. and Kabla, I. (1998). What percentage of innovations are patented? Empirical estimates for European firms. *Research Policy*, 27(2):127–141.

Audretsch, D. B. and Feldman, M. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, 86(4):253–273.

Aunger, R. (2010). Types of technology. *Technological Forecasting and Social Change*, 77(5):762–782.

Balland, P. A. and Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, 93(1):1–23.

[15]The results for additional years can be downloaded from http://www.tombroekel.de/research/technological-complexity/.

Bonchev, D. and Buck, G. (2005). Quantitative measures of network complexity. In Bonchev, D. and Rouvrary, D., editors, *Complexity in chemistry biology and ecology*, chapter 5. Springer Verlag, New York.

Breschi, S. and Lenzi, C. (2011). Net City: How co-invention networks shape inventive productivity in U.S. cities. *KITeS Seminarpapers*, pages 1–32.

Camagni, R. (1991). Local "milieu", uncertainty and innovation networks: towards a new dynamic theory of economic space. In Camagni, R., editor, *Innovation Networks: Spatial Perspectives*, pages 121–142. Belhaven Stress. London, UK and New York, USA.

Carbonell, P. and Rodriguez, A. I. (2006). Designing teams for speedy product development: The moderating effect of technological complexity. *Journal of Business Research*, 59(2):225–232.

Coad, A. and Rao, R. (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research Policy*, 37(4):633–648.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1):128–152.

Cooke, P. (1992). Regional Innovation Systems: Competitive Regulation in the New Europe. *GeoForum*, 23:356–382.

Dalmazzo, A. (2002). Technological complexity, wage differentials and unemployment. *Scandinavian Journal of Economics*, 104(4):515–530.

Dehmer, M., Barbarini, N., Varmuza, K., and Graber, A. (2009). A large scale analysis of information-theoretic network complexity measures using chemical structures. *PLOS ONE*, 4(12):20–26.

Dehmer, M. and Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences*, 181(1):57–78.

Emmert-Streib, F. and Dehmer, M. (2012). Exploring statistical and population aspects of network complexity. *PLoS ONE*, 7(5).

EPO, JPO, and USPTO (2007). *Trilateral statistical report*. European patent office, Munich.

Fai, F. and Von Tunzelmann, N. (2001). Industry-specific competencies and converging technological systems: Evidence from patents. *Structural Change and Economic Dynamics*, 12(2):141–170.

Fleming, L. and Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7):1019–1039.

Florida, R. (1995). Toward the Learning Region. *Futures*, 27:527–536.

Goldreich, O., Ostrovsky, R., and Petrank, E. (1998). Computational complexity and knowledge complexity. *SIAM Journal on Computing*, 27(4):1116–1141.

Goldreich, O. and Petrank, E. (1999). Quantifying Knowledge Complexity. *Comput. complex*, 8(November):50–98.

Griffin, A. (1997). The Effect of Project and Process Characteristics on Product Development Cycle Time. *Journal of Marketing Research*, 34(1):24–35.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28:1661–1701.

Hägerstrand, T. (1967). *Innovation diffusion as a spatial process*. University of Chicago Press, Chicago.

Hargadon, A. (2003). *How breakthroughs happen - The surprising truth about how companies innovate*. Harvard Businesss School Press, Boston.

Hidalgo, A. and Hausmann, R. (2009). The building blocks of economic complexity. *PNAS*, 106(26):10570–10575.

Hidalgo, C. A. (2015). *Why Information Grows: The Evolution of Order, from Atoms to Economies*. Basic Books, New York.

Howitt, P. (1999). Steady Endogenous Growth with Population and R . & D . Inputs Growing. *Journal of Political Economy*, 107(4):715–730.

Jaffe, A. B. (1989). Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy*, 18(2):87–97.

Kim, B. W. (2015). Economic Growth: Education vs. Research. *Journal of Global Economics*, 03(04).

Kim, S. and Anand, J. J. (2018). Knowledge complexity and the performance of inter-unit knowledge replication structures. *Strategic Management Journal*, 39(7):1959–1989.

Kogut, B. and Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3):383–397.

Madsen, J. B. (2007). Are there diminishing returns to R&D? *Economics Letters*, 95(2):161–166.

Mcnerney, J., Farmer, J. D., Redner, S., and Trancik, J. E. (2011). Role of design complexity in technology improvement. *PNAS*, 108(22):9008–9013.

Mendonça, S. (2009). Brave old world: Accounting for 'high-tech' knowledge in 'low-tech' industries. *Research Policy*, 38(3):470–482.

Milo, R., Stumpf, M. P., and Stark, J. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827.

Nelson, R. R. and Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *American Economic Review*, 72(1):114–132.

Pavitt, K. (1998). Technologies, products and organization in the innovating firm: what Adam Smith tells us and Joseph Schumpeter doesn't. *Industrial and Corporate Change*, 7:433–452.

Pintea, M. and Thompson, P. (2007). Technological complexity and economic growth. *Review of Economic Dynamics*, 10(2):276–293.

Prencipe, A. (2000). Breadth and depth of technological capabilities in CoPS: the case of the aircraft engine control system. *Research Policy*, 29:895–911.

Rivkin, J. W. (2000). Imitation of Complex Strategies. *Management Science*, 46(6):824–844.

Rogers, E. M. (1995). *The Diffusion of Innovation, 4th ed.* Free Press, New York.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98:71–102.

Sáenz, J., Aramburu, N., and Rivera, O. (2009). Knowledge sharing and innovation performance: A comparison between high-tech and low-tech companies. *Journal of Intellectual Capital*, 10(1):22–36.

Schmoch, U., Laville, F., Patel, P., and Frietsch, R. (2003). Linking technology areas to industrial sectors. *Final Report to the European Commission, DG Research, Karlsruhe, Paris, Brighton*.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(July 1928):379–423.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the american philosphical society*, 106(6):467–482.

Singh, K. (1997). The impact of technological complexity and interfirm cooperation on business survival. *Academy of Management Journal*, 40(2):339–367.

Sorenson, O. (2005). Social Networks, Informational Complexity and Industrial Geography. In Fornahl, D., Zellner, C., and Audretsch, D., editors, *The Role of Labour Mobility and Informal Networks for Knowledge Transfer*, chapter 5, pages 79–96. Springer Science+Business Media Inc., Boston.

Sorenson, O. and Fleming, L. (2004). Science and the diffusion of knowledge. *Research Policy*, 33(10):1615–1634.

Sorenson, O., Rivkin, J. W., and Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, 35(7):994–1017.

Steward, D. V. (1981). The design structure system: A method for managing the design of complex systems. *IEEE Engineering Management*, 28:71–74.

Teece, D. J. (1977). Technology Transfer by Multinational Firms : The Resource Cost of Transferring Technological Know-How. *Economic Journal*, 87:242–261.

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69(1):17–20.

Wikipedia (2017). Source lines of code.

Ye, F. Y. (2017). Measuring knowledge: A quantitative approach to knowledge theory. In *Scientific Metrics: Towards Analytical and Quantitative Sciences. Understanding Complex Systems*, pages 155–162. Springer, Singapore.

Zander, U. and Kogut, B. (1995). Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test. *Organization Science*, 6(1):76–91.

Appendix

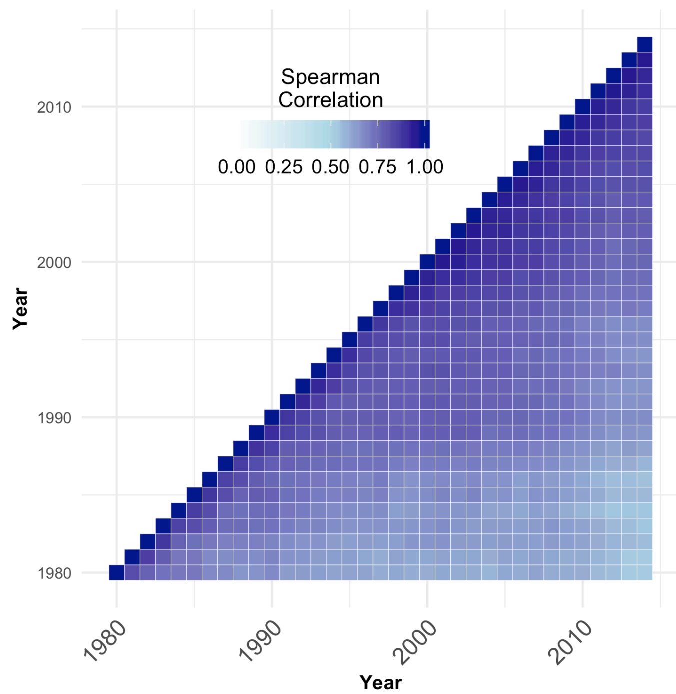


Figure 9: Temporal correlation of *structural diversity*.

Table 3: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Structural diversity	23,104	7.46	2.93	0.00	5.71	9.63	14.98
Patents	23,312	766.18	1,989.02	1.00	65.00	653.25	33,414.00
High-tech	23,580	0.04	0.21	0	0	0	1
Median age	23,473	6.66	3.62	0.00	3.76	9.13	25.34
Inventors per patent	23,312	1.95	0.51	0.50	1.59	2.25	12.00
Spatial Gini	23,190	0.90	0.06	0.69	0.86	0.95	1.00
CPCs per patent	22,932	4.64	2.83	1.00	3.00	6.00	68.00
FS.Modular	22,607	0.99	0.53	0.00	0.68	1.17	9.43
HH.NUTS2	23,312	88.02	17.17	0.00	77.00	99.98	100.00

Table 4: (Rank) correlation Matrix

	Structural	Patents	High tech	Median age	Inventors per patent	Spatial Gini	CPCs per patent	FS. Modular
Patents	0.66							
High-tech	0.07	0.09						
Median age	0.27	0.1	-0.05					
Inventors per patent	0.54	0.32	0.09	0.32				
Spatial Gini	-0.47	-0.8	0	-0.06	-0.15			
CPCs per patent	0.7	0.31	0.01	0.34	0.52	-0.24		
FS.Modular	-0.18	0.05	0.09	0.18	0.15	0.12	-0.4	
HH.NUTS2	-0.3	-0.21	-0.01	-0.49	-0.26	0.17	-0.29	-0.09

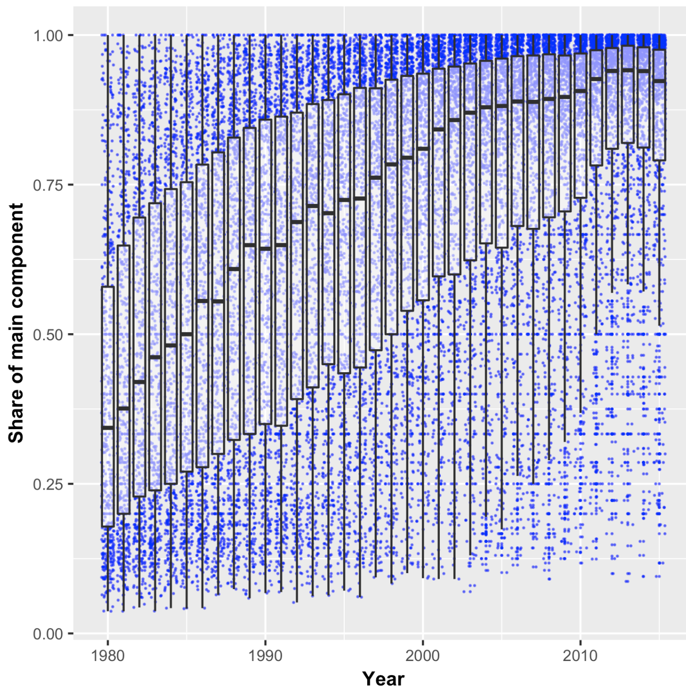


Figure 10: Share of main component

Table 5: Characteristics of *structural diversity*, 2001-2015

	Dependent variable:							
	Structural diversity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Patents)	0.969***	0.967***	0.821***	0.716***	0.757***		1.086***	0.749***
	(0.037)	(0.037)	(0.041)	(0.037)	(0.054)		(0.054)	(0.041)
High-tech		0.107	−0.011	−0.245	−0.266	0.966***		−0.048
		(0.249)	(0.237)	(0.215)	(0.215)	(0.300)		(0.178)
Log(Median age + 1)			−2.633***	−2.666***	−2.613***			−1.845***
			(0.289)	(0.255)	(0.258)			(0.217)
Log(Inventors per patent)				3.130***	3.095***			1.128***
				(0.205)	(0.205)			(0.169)
Log(Spatial Gini)					1.117	−14.831***	3.433***	2.671***
					(1.013)	(1.029)	(1.142)	(0.690)
Log(CPCs per patent)								2.130***
								(0.083)
adj. R2	0.456	0.456	0.498	0.577	0.575	0.233	0.459	0.702
n	643	643	643	643	643	643	643	643
T	15	15	15	15	15	15	15	15
N	9,433	9,433	9,433	9,433	9,425	9,425	9,425	9,407
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Unbalanced panel regression, robust standard errors and p-values. *p<0.1; **p<0.05; ***p<0.01

Table 6: Technologies and *structural diversity* in 2014

Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*
1	B60L	3,098	13.356	61	B60K	3,438	11.782	121	F03B	488	11.175
2	Y04S	1,233	13.227	62	D04H	747	11.766	122	H02P	1,881	11.174
3	H04W	22,610	12.922	63	B60M	176	11.760	123	A43B	886	11.171
4	B60W	2,551	12.731	64	A23V	2,263	11.725	124	H04R	2,791	11.164
5	C10N	788	12.668	65	B42D	684	11.725	125	C25D	985	11.157
6	F17C	706	12.661	66	C10G	1,392	11.704	126	C08L	6,697	11.143
7	H03F	826	12.631	67	B23K	3,419	11.699	127	G06Q	8,663	11.143
8	Y02C	614	12.617	68	F05B	1,905	11.696	128	B29B	1,078	11.136
9	B33Y	1,211	12.566	69	C12Y	2,347	11.684	129	H01G	1,152	11.134
10	A61H	926	12.518	70	H04N	13,841	11.670	130	H05K	4,911	11.116
11	C21D	1,607	12.516	71	B29L	4,310	11.631	131	B60B	698	11.108
12	F02D	2,213	12.504	72	H04S	769	11.612	132	F02B	1,833	11.106
13	F01N	2,222	12.454	73	B21B	548	11.611	133	F01L	691	11.075
14	C01P	1,504	12.427	74	C10L	948	11.606	134	C12M	1,281	11.067
15	C12Q	4,428	12.413	75	F21W	624	11.600	135	E05Y	1,315	11.060
16	F21K	1,033	12.396	76	C05F	232	11.596	136	B41M	1,021	11.060
17	F05D	5,581	12.389	77	C11D	1,720	11.553	137	F23G	303	11.059
18	H02M	3,240	12.362	78	C12P	2,419	11.546	138	A61N	3,343	11.058
19	F01D	5,568	12.349	79	C08J	4,045	11.528	139	B44C	383	11.018
20	F21S	1,957	12.341	80	C13K	172	11.526	140	A23C	497	11.018
21	B22F	1,662	12.286	81	G10L	2,090	11.514	141	H04Q	1,051	11.003
22	A61Q	3,657	12.282	82	C22F	599	11.512	142	C09B	607	11.002
23	G06T	7,184	12.271	83	A23Y	154	11.493	143	A61L	3,766	10.995
24	C01G	768	12.267	84	A43C	249	11.491	144	A21D	333	10.980
25	H02J	5,532	12.256	85	D06N	228	11.481	145	C01F	393	10.966
26	B32B	6,486	12.254	86	B29K	4,060	11.479	146	G07C	1,416	10.956
27	C09C	583	12.222	87	B82Y	1,555	11.479	147	G06K	6,900	10.950
28	C10M	917	12.208	88	D03D	473	11.451	148	F28D	2,030	10.917
29	F21Y	2,240	12.175	89	A23D	287	11.443	149	G01C	2,759	10.901
30	G09G	2,743	12.171	90	A61B	17,651	11.438	150	H04H	577	10.900
31	D07B	234	12.166	91	F25B	2,098	11.434	151	E04F	1,145	10.889
32	B01L	1,941	12.159	92	B29C	8,927	11.425	152	E05F	999	10.883
33	C10J	262	12.139	93	F24J	861	11.423	153	H01Q	2,440	10.867
34	A61K	22,848	12.133	94	A01N	2,830	11.413	154	D10B	839	10.867
35	C22C	3,006	12.122	95	C08K	5,155	11.384	155	C08G	5,480	10.860
36	F23R	1,040	12.078	96	B60G	940	11.371	156	C07B	828	10.854
37	C10K	226	12.049	97	C25B	707	11.365	157	B60Q	1,176	10.845
38	H01M	7,041	12.041	98	F04D	2,630	11.346	158	H04B	8,740	10.843
39	A63B	1,276	12.007	99	H01S	980	11.331	159	C09D	5,282	10.842
40	F03D	2,108	11.996	100	A23P	431	11.330	160	C21B	218	10.835
41	C07K	7,229	11.980	101	C10B	281	11.323	161	B05D	1,878	10.823
42	C09J	2,366	11.972	102	D21H	1,074	11.309	162	A45D	1,080	10.817
43	C07D	8,552	11.965	103	G08C	706	11.302	163	A24F	742	10.817
44	B60Y	1,263	11.958	104	F02P	335	11.298	164	G06F	32,551	10.811
45	H04J	1,979	11.954	105	Y02T	10,645	11.283	165	A23G	670	10.803
46	H04L	27,170	11.921	106	C11B	461	11.274	166	C12R	320	10.801
47	A23K	757	11.913	107	H04M	4,799	11.268	167	G02C	972	10.799

48	F02N	490	11.908	108	F23J	340	11.268	168	D01D	460	10.798
49	H01L	12,946	11.907	109	B62M	588	11.258	169	G06N	1,121	10.793
50	F25J	319	11.901	110	F15B	1,187	11.246	170	F02M	2,530	10.790
51	C02F	2,293	11.871	111	C07C	5,460	11.244	171	C01B	2,858	10.783
52	E02F	1,157	11.856	112	A61M	8,004	11.231	172	B60C	2,324	10.780
53	F02C	2,991	11.852	113	F15C	8	11.205	173	C30B	565	10.765
54	C04B	2,787	11.849	114	B22D	977	11.200	174	H02K	3,912	10.765
55	F21V	2,975	11.827	115	B61C	165	11.194	175	D04B	464	10.734
56	F02K	880	11.813	116	B28B	584	11.183	176	B62K	978	10.724
57	B61L	499	11.812	117	B01J	5,367	11.181	177	F28F	1,958	10.724
58	A23L	3,142	11.793	118	B60T	1,665	11.179	178	A23J	194	10.723
59	C12N	8,072	11.792	119	F01K	868	11.178	179	H02H	1,323	10.709
60	G08G	1,745	11.787	120	G03H	194	11.176	180	E05D	921	10.703

Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*
181	A01H	331	10.700	242	C07H	810	10.190	303	C22B	575	9.736
182	F21L	117	10.683	243	F16C	2,624	10.187	304	B64D	2,669	9.734
183	B06B	468	10.682	244	B29D	2,066	10.184	305	B23P	1,458	9.730
184	F24D	891	10.682	245	F04B	1,847	10.183	306	A63F	812	9.704
185	B31B	442	10.627	246	C05B	103	10.173	307	A46D	119	9.703
186	D21C	340	10.621	247	G03F	1,097	10.157	308	F16L	3,062	9.699
187	E04C	842	10.618	248	A41D	745	10.139	309	B25F	548	9.663
188	G04G	280	10.616	249	A47L	1,911	10.135	310	E05C	587	9.660
189	G01J	1,550	10.614	250	A61F	6,780	10.134	311	B61H	94	9.659
190	C08C	267	10.609	251	B25J	1,585	10.125	312	G05G	404	9.636
191	G11C	1,104	10.583	252	A45F	246	10.121	313	C09K	3,399	9.635
192	A23F	248	10.573	253	B27D	99	10.121	314	B24D	301	9.635
193	G09C	262	10.553	254	B41C	143	10.113	315	B21C	392	9.626
194	F04C	1,615	10.548	255	F25C	192	10.103	316	E04D	502	9.620
195	G07D	605	10.532	256	G01T	702	10.095	317	F16D	2,659	9.614
196	F01C	345	10.531	257	G08B	1,970	10.090	318	A61G	1,017	9.612
197	E05B	1,576	10.529	258	A45C	434	10.082	319	B23C	481	9.601
198	G03B	1,475	10.524	259	B03C	310	10.077	320	H04K	99	9.600
199	C21C	267	10.520	260	G01S	4,267	10.072	321	H03L	306	9.566
200	F23L	250	10.513	261	C05G	201	10.056	322	H01P	550	9.557
201	F01P	492	10.504	262	C05C	102	10.042	323	B42C	106	9.546
202	D06F	1,805	10.492	263	G05B	4,570	10.024	324	F23M	186	9.524
203	F24F	2,029	10.491	264	B60R	3,738	10.023	325	A24D	362	9.518
204	F16M	822	10.475	265	A41B	176	10.012	326	B65H	1,930	9.511
205	H03G	406	10.467	266	H01F	2,431	9.992	327	B62B	472	9.509
206	A61C	1,588	10.449	267	B27K	96	9.989	328	B65F	245	9.508
207	C08H	248	10.431	268	C03B	807	9.987	329	B03D	100	9.496
208	D01F	696	10.430	269	Y02E	11,568	9.980	330	B60H	793	9.492
209	B64C	2,314	10.420	270	D05C	57	9.978	331	B63G	189	9.485
210	D02G	287	10.416	271	F23D	610	9.962	332	C05D	130	9.484
211	C07J	192	10.383	272	B26D	813	9.959	333	G11B	1,080	9.474
212	C07F	2,073	10.383	273	G07F	1,022	9.946	334	F16B	2,539	9.473
213	B04B	239	10.382	274	B21J	380	9.943	335	F23C	414	9.469
214	G02F	2,326	10.382	275	G06M	47	9.935	336	H02N	386	9.449
215	B04C	126	10.374	276	B81C	386	9.929	337	C40B	115	9.449
216	H01R	3,558	10.374	277	C12C	128	9.926	338	B01F	1,588	9.448
217	B81B	432	10.362	278	G01N	15,372	9.926	339	B64G	313	9.441
218	F25D	1,619	10.360	279	H05B	4,271	9.921	340	D21B	88	9.439
219	B01D	6,664	10.359	280	Y02B	7,205	9.917	341	G07B	278	9.435
220	A46B	541	10.355	281	E21B	3,963	9.913	342	G01B	2,531	9.425
221	H03M	1,123	10.336	282	E04B	1,625	9.905	343	B64F	500	9.421
222	G09B	1,095	10.317	283	A61J	1,091	9.887	344	Y02P	6,950	9.416
223	B65B	2,495	10.309	284	C07G	91	9.884	345	C23G	151	9.415
224	C23C	3,703	10.295	285	G09F	879	9.877	346	B62J	661	9.408
225	A23B	312	10.276	286	F27D	713	9.876	347	C01D	126	9.404
226	C03C	1,468	10.271	287	G10K	759	9.867	348	F01M	474	9.403
227	H01B	2,038	10.269	288	B27N	132	9.854	349	G21D	156	9.400
228	B25D	242	10.268	289	G01V	2,028	9.852	350	B23H	195	9.398

229	G10H	258	10.264	290	G01R	4, 554	9.829	351	F05C	149	9.388
230	B41J	2, 388	10.260	291	B67C	330	9.824	352	Y10T	17, 313	9.388
231	H02S	755	10.254	292	G07G	175	9.811	353	D04C	94	9.380
232	Y02W	1, 474	10.248	293	H05H	475	9.805	354	A01D	1, 027	9.374
233	C08F	3, 794	10.243	294	F27B	543	9.802	355	B63H	537	9.371
234	F23N	406	10.240	295	G05F	575	9.801	356	E03C	472	9.352
235	C08B	678	10.239	296	G02B	9, 406	9.795	357	C01C	152	9.313
236	F24H	656	10.235	297	F16H	3, 579	9.781	358	H01J	1, 914	9.306
237	B09B	263	10.229	298	B65D	5, 994	9.770	359	G21F	319	9.297
238	A24B	290	10.222	299	G21K	347	9.760	360	F03G	370	9.288
239	D06M	641	10.211	300	F02G	181	9.758	361	D06P	212	9.283
240	G21C	449	10.206	301	F16N	249	9.742	362	B08B	1, 083	9.278
241	B60N	1, 353	10.199	302	A43D	139	9.740	363	B67D	583	9.271

Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*
364	H05G	168	9.268	425	F16J	1,351	8.708	486	G01Q	113	8.092
365	B09C	91	9.267	426	A47C	973	8.704	487	G01G	454	8.091
366	B62D	3,054	9.250	427	B07B	243	8.686	488	C23D	25	8.080
367	B61K	96	9.238	428	F01B	171	8.682	489	F27M	12	8.064
368	H01C	260	9.219	429	B68F	2	8.675	490	G04R	66	8.062
369	B42B	52	9.218	430	A21C	182	8.664	491	B25C	196	8.060
370	B23B	976	9.218	431	G21Y	45	8.645	492	E05G	61	8.042
371	B66B	826	9.216	432	B62H	127	8.643	493	B66C	742	8.040
372	C12G	114	9.215	433	C25F	111	8.635	494	E03D	326	8.030
373	B02C	759	9.185	434	A62B	467	8.630	495	A45B	95	8.026
374	B43K	138	9.182	435	F24C	1,079	8.628	496	A44C	251	7.996
375	G01P	892	9.176	436	B24C	211	8.626	497	H02B	462	7.988
376	C25C	179	9.166	437	B21D	1,393	8.591	498	A41H	47	7.983
377	A42B	294	9.155	438	B63J	151	8.588	499	D21F	321	7.981
378	B22C	374	9.145	439	A01C	514	8.571	500	A47F	528	7.980
379	B60D	261	9.134	440	A63C	303	8.565	501	B25B	857	7.969
380	F22B	456	9.115	441	F23B	92	8.548	502	F28C	72	7.912
381	B62L	105	9.113	442	B65G	2,574	8.543	503	A23N	150	7.910
382	B82B	38	9.112	443	F28B	71	8.541	504	H03C	43	7.904
383	E06B	1,811	9.107	444	B60J	803	8.529	505	A47G	638	7.890
384	B05B	2,012	9.101	445	B44B	66	8.527	506	C06B	109	7.866
385	F02F	467	9.088	446	F26B	560	8.512	507	F41C	91	7.866
386	E02B	403	9.056	447	F41G	283	8.501	508	E01C	566	7.852
387	B65C	286	9.046	448	D02J	68	8.492	509	F42B	447	7.848
388	F23K	189	9.030	449	G01H	370	8.483	510	B21H	97	7.844
389	A44B	308	9.011	450	A61D	173	8.470	511	A01F	438	7.841
390	A01K	1,497	8.998	451	G01F	1,661	8.462	512	F16G	336	7.835
391	B26F	316	8.980	452	B60S	643	8.446	513	F28G	80	7.798
392	H03K	1,756	8.961	453	A41F	101	8.441	514	A47D	135	7.783
393	F16K	2,763	8.949	454	F15D	128	8.438	515	A63H	344	7.735
394	A22C	351	8.947	455	E21C	172	8.428	516	D01G	138	7.733
395	B64B	47	8.937	456	G01W	118	8.406	517	B01B	18	7.718
396	Y10S	1,752	8.934	457	A01B	606	8.404	518	B66D	223	7.716
397	A47J	2,161	8.933	458	D06B	127	8.404	519	F16P	168	7.708
398	G01K	917	8.911	459	D06C	100	8.356	520	B60P	593	7.707
399	C09G	126	8.908	460	H03D	116	8.338	521	B28D	303	7.703
400	D06Q	25	8.895	461	B31D	128	8.333	522	F41H	416	7.689
401	B05C	713	8.883	462	E04H	1,033	8.331	523	A01M	395	7.660
402	G05D	2,300	8.871	463	E02D	689	8.329	524	A63G	124	7.655
403	A44D	32	8.870	464	B21K	249	8.321	525	B23D	583	7.630
404	B27M	105	8.866	465	B28C	97	8.319	526	F22G	47	7.590
405	H01H	2,499	8.863	466	C11C	198	8.317	527	F23Q	144	7.585
406	H01T	315	8.839	467	A01J	155	8.302	528	B41K	41	7.576
407	H02G	1,550	8.815	468	F16F	1,767	8.301	529	G21B	61	7.574
408	B41F	682	8.801	469	G01M	2,017	8.259	530	D01H	235	7.572
409	B44F	96	8.796	470	B68G	49	8.257	531	F22D	53	7.549
410	G04B	654	8.782	471	C12F	27	8.244	532	F24B	92	7.547
411	C23F	368	8.761	472	C10C	31	8.242	533	E03B	195	7.528

412	F03C	105	8.754	473	A47K	683	8.231	534	B24B	856	7.509
413	B07C	297	8.748	474	B23F	126	8.227	535	E01D	124	7.507
414	F17D	191	8.742	475	A41C	92	8.218	536	B61F	192	7.460
415	F42D	89	8.741	476	A47B	1,183	8.208	537	A24C	279	7.393
416	A62D	110	8.739	477	H03B	140	8.191	538	D01B	8	7.367
417	G01D	1,931	8.736	478	G06E	17	8.173	539	G04C	171	7.321
418	B41P	175	8.735	479	G10G	28	8.164	540	E03F	324	7.318
419	H03H	508	8.735	480	B30B	534	8.162	541	A21B	137	7.316
420	B66F	599	8.732	481	G01L	1,553	8.146	542	B43M	29	7.316
421	G03G	1,548	8.730	482	B61D	440	8.146	543	B25G	131	7.288
422	A62C	454	8.725	483	B41N	137	8.135	544	F04F	156	7.276
423	B03B	91	8.717	484	B23Q	1,019	8.114	545	E04G	570	7.257
424	B63B	1,227	8.709	485	B31F	187	8.096	546	A01G	1,027	7.251

Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*	Rank	CPC	Patents	*Structural diversity*
547	B21F	148	7.241	584	F41F	42	6.514	621	G10F	15	4.356
548	F03H	30	7.227	585	F42C	97	6.406	622	A42C	18	4.141
549	C13B	38	7.215	586	B67B	136	6.350	623	A63D	17	4.079
550	D05B	130	7.183	587	B27G	84	6.336	624	G12B	10	4.042
551	G21G	61	7.141	588	B60F	37	6.268	625	B42P	19	3.794
552	D21D	76	7.111	589	E06C	102	6.193	626	B68C	33	3.329
553	B42F	45	7.100	590	C06D	47	6.137	627	B60V	10	3.156
554	C14C	44	7.099	591	C14B	50	6.134	628	B68B	13	3.064
555	C12H	58	7.064	592	F23H	26	6.122	629	H01K	28	2.682
556	B61B	147	7.039	593	D03C	29	6.108	630	G09D	2	2.679
557	B44D	40	7.022	594	E01B	290	6.069	631	C12J	5	2.209
558	E01F	284	7.018	595	B31C	18	6.033	632	H05C	6	2.123
559	D21J	32	7.009	596	E21D	154	5.993	633	B02B	5	1.171
560	H03J	54	6.992	597	G04D	65	5.973	634	B41G	2	0
561	B63C	201	6.980	598	A22B	83	5.933	635	B61J	8	0
562	F41A	351	6.956	599	B27C	41	5.917	636	C06F	2	0
563	B27B	262	6.902	600	E21F	58	5.901	637	C09H	1	0
564	D06H	53	6.895	601	B21G	11	5.813	638	C10F	2	0
565	G06G	27	6.889	602	B27L	40	5.785	639	C10H	2	0
566	D21G	117	6.877	603	A41G	40	5.778	640	C12L	1	0
567	B41D	9	6.851	604	D06J	2	5.760	641	D04G	5	0
568	B27J	5	6.849	605	A63J	68	5.595	642	D05D	11	0
569	B43L	45	6.837	606	H05F	38	5.537	643	D06G	3	0
570	A47H	64	6.830	607	B27H	2	5.326	644	E02C	1	0
571	D06L	44	6.810	608	A01L	30	5.280	645	G06C	2	0
572	D01C	18	6.806	609	F16S	9	5.260	646	G06J	1	0
573	G03C	31	6.805	610	B25H	201	5.226	647	B41B	–	–
574	G04F	141	6.754	611	F41J	53	5.214	648	B62C	–	–
575	D03J	25	6.729	612	D04D	12	5.182	649	F17B	–	–
576	C09F	23	6.706	613	F16T	21	5.090	650	F21H	–	–
577	B26B	406	6.657	614	G10D	111	5.008	651	G03D	–	–
578	F41B	84	6.622	615	E01H	148	4.986	652	G06D	–	–
579	B61G	68	6.595	616	G21H	14	4.957	653	G10B	1	–
580	B23G	49	6.587	617	A63K	14	4.723	654	G21J	–	–
581	B41L	10	6.570	618	B27F	28	4.623	655	H04T	–	–
582	C06C	26	6.563	619	D02H	6	4.513				
583	G10C	33	6.526	620	B21L	14	4.431				