

Technology cycles and the evolution of the knowledge base complexity since the 1980s

* *Vanessa de Lima Avanci* 

Universidade Federal Fluminense (UFF), Niterói (RJ), Brasil

E-mail: vanessa.avanci@gmail.com

** *Ana Urraca Ruiz* 

Universidade Federal Fluminense (UFF), Niterói (RJ), Brasil

E-mail: anauracaruz@gmail.com

RECEIVED: 26 MAY 2019 REVISED VERSION: 10 MARCH 2020 ACCEPTED: 09 NOVEMBER 2020

ABSTRACT

Knowledge bases are complex systems of integrated technological knowledge that represent solutions to specific problems for the state of the art of knowledge in a given period. Knowledge bases evolve with technological cycles. In the last 30 years, we identify two technological waves. The first one started with the seven technological paradigms in the 1980s (microelectronics, computers, telecommunications, audiovisual, new materials, semiconductors and biotechnology); the second one initiated with the so-called Key Enabling Technologies (KETs - nanotechnology, micro and nanoelectronics, industrial biotechnology, photonics, advanced materials and advanced manufacturing) from the 2000. This paper analyzes the evolution of the properties and complexity of the world knowledge base over 1978-2016. Using patent data and network analysis, the work calculates indicators for variety, coherence, cognitive distance and convergence of the knowledge base. The results confirm that the technological paradigms of the 1980s are associated with an increase in the diversification and complexity of the knowledge base through an outward convergence, that is, with not related technologies - inside the same paradigm. The arrival of the 2000 micro-paradigms reveals a retraction of the knowledge base that evolves towards more concentrated paths over the trajectories previously established.

KEYWORDS | Innovation; knowledge base; Technological paradigms; Convergence; Networks.

1. Introduction

A knowledge base for a technology or a group of technologies is a representation of interconnected knowledge subunits that reveal the 'state of the art' of available methods, processes, skills and techniques at any given time. These subunits, or pieces of knowledge, interrelate in a highly complex way to respond to specific functionalities (SAHAL, 1981; ARTHUR, 2009; SAVIOTTI, 2009). A knowledge base evolves cyclically along with technical progress, alternating emergence and maturation phases of new scientific-technical paradigms (SAVIOTTI, 2009). When a new paradigm emerges, new pieces of knowledge or new interconnections emerge. As technological standards set in, uncertainties diminish, and the various directions of technical progress define themselves in technological trajectories or clusters of possible technological directions (DOSI, 1984). The development of a technological trajectory leads to incremental changes in the knowledge base and the base is transformed by the recombination between pre-existing parts. This transformation process happens throughout the innovation-diffusion process (FLEMING; SORENSON, 2001; ARTHUR, 2009).

Knowledge bases have four properties that scale their complexity: variety, convergence, coherence and cognitive distance (KRAFFT; QUATRARO; SAVIOTTI 2011). *Variety* refers to the extent of base diversification in terms of the quantity of elements and the multiplicity of combinations formed between them. When new pieces of knowledge emerge with new technological paradigms, they combine with other previously existing pieces and can establish strong complementarities over time. The more often different pieces of knowledge combine, the greater the degree of similarity between them; and the greater the number of interconnections between similar or dissimilar technologies, the greater the degree of *convergence*. *Coherence* refers to the way and the intensity by which the pieces are integrated in the knowledge base as a whole (NESTA; SAVIOTTI, 2005). Usually, the interrelationship between pieces of knowledge follows specific patterns according to the functionality of the technology. *Cognitive distance* refers to the degree of dissimilarity between pieces of knowledge. Measures of cognitive distance try to identify, on the one hand, discontinuities in the base, that is, the emergence of new pieces that should appear to be poorly connected. On the other hand, a reduction in the cognitive distance between pieces would point to a process of cumulative knowledge development, that is, a maturation phase (KRAFFT; QUATRARO; SAVIOTTI, 2011).

The 1980s were characterized by the emergence of seven technological paradigms: microelectronics, computers, telecommunications, audiovisual, new materials, semiconductors and biotechnology (FREEMAN; PEREZ, 1988), and the 2000s by the *Key Enabling Technologies* (KETs), which are nanotechnology, micro and nanoelectronics, industrial biotechnology, photonics, advanced materials and advanced manufacturing (VAN DE VELDE *et al.*, 2015). Assuming that new paradigms alter the properties of the global knowledge base, this work aims to analyze the evolution of the global knowledge base complexity as the technological trajectories of the paradigms were shaping during the 1980-2016 period. To do that, the paper uses usual network analysis and patent data to create indicators that measure the changes registered in the properties of the global knowledge base: variety, coherence, cognitive distance and convergence. The main contribution of the article, in this sense, is that it is a complexity analysis of the global knowledge base by periods, which not only contemplates the interaction between pieces of knowledge related to specific technologies, but also in terms of integrity in its completeness, that is, considering still the interactions among paradigms to define common trajectories.

Besides this introduction, the paper first presents how the knowledge base should evolve over technological cycles given its properties. Next, the article sets out the methodology used; concretely, it discusses the feasibility of using patents to elaborate knowledge bases, as well as the description of the indicators used to measure the properties of a knowledge base: variety, coherence, cognitive distance and convergence. The article ends discussing the main results found.

2. Global knowledge base: structure and evolution

Knowledge is an abstract structure that relates to different subunits (KRAFFT; QUATRARO; SAVIOTTI, 2011). A knowledge base referred to a technology system is a complex system of pieces of technological knowledge that interdependently and non-randomly combine to solve specific problems – functionalities- over a given period (for example electric car, heat resistant materials, etc.). Techno-scientific paradigms, as well as technological micro-paradigms, define knowledge bases specific to the functionalities the paradigm is associated with.

The knowledge base evolves with the technological cycle (SAVIOTTI, 2009). In the initial phase of a new paradigm, new pieces of knowledge can emerge as long as some of the old ones can disappear. Solving specific problems within a given set of knowledge will be able to occur in isolation from the others. In addition,

new functionalities emerge and the possibilities for recombination of old pieces of knowledge with the new ones will lead to greater variety. Consequently, the knowledge base becomes less coherent (NESTA; SAVIOTTI, 2005). Those effects characterize the evolution of technical progress by discontinuity, that is, the natural process of creative destruction in the Schumpeterian sense. In mature phases, technical progress evolves along specific paths among a diversity of possible directions. Occasionally, several trajectories may combine, while others may cease to exist [for example, analog cellular technologies] (VERSPAGEN, 2007). The success of a trajectory or the possibility of offering greater technological opportunities will lead to the associated pieces of knowledge generating interrelationships or new combinations at a faster rate in relative terms. The increase or decrease in the frequency with which a given relationship occurs and the creation/suppression of links between pieces of knowledge characterize the continuity of the technical progress of this phase (KRAFFT; QUATRARO; SAVIOTTI, 2011). Greater interdependence means greater coherence if complementary technologies grow at the same pace and direction. However, new combinations of functionalities do not use to occur at the same rate as completely new pieces of knowledge are generated (SAVIOTTI, 2009). Instead, greater interdependence introduces tradeoffs in the system, this is, the advancement in certain technological domains only happen if their complementary domains also advance (SAHAL, 1985).

New connections in the base can combine similar or dissimilar technological domains (NESTA; SAVIOTTI, 2005). ‘Similarity’ refers to proximity, that is, to pieces of knowledge that belong to the same technological domain - as ‘Dissimilarity’ refers to pieces of knowledge that belong to different technological domains in terms of application or use for a given level of aggregation. For example, antibiotics and vaccines are similar as both solve health problems, but they are dissimilar to the extent that the first belongs to ‘healing medicine’ and the second to ‘preventive medicine’. When connections occur between dissimilar technological domains in their functionality there are technological *convergence* (NESTA; SAVIOTTI, 2005). Empirical studies have observed increased combinations between pieces of knowledge that emerged with new paradigms from the 2000s and those that previously existed (JOO; KIM, 2010; KRAFFT; QUATRARO; SAVIOTTI, 2011; KIM; CHO; KIM, 2014). More sophisticated measures of the degree of interrelationship between pieces of knowledge allow us to observe changes in the degree of similarity between areas of knowledge over time (JOO; KIM, 2010; YAN; LUO, 2017). These works revealed that there was quite a lot new connections that belong to dissimilar areas

of knowledge, particularly in technologies such as microelectronics, semiconductors and informatics (JOO; KIM, 2010; HUENTELER *et al.*, 2016).

Table 1 summarizes the hypotheses about how complexity and convergence should evolve along the technological cycle considering the four mentioned properties of the knowledge base: variety, coherence, cognitive distance and convergence. Discontinuity characterizes the initial stage; new pieces of knowledge and new combinations emerge. If the new elements are little connected and relatively distant from the others, variety will increase and coherence will decrease. As the technological cycle matures, the knowledge base evolves by continuity; new connections and complementarities will emerge between old and new pieces of knowledge. The knowledge base becomes more complex and integrated; therefore, coherence and convergence will rise. Technical progress will evolve around specific trajectories; therefore, variety and cognitive distance will reduce.

TABLE 1
Assumptions on the evolution of knowledge base properties

	Properties of the knowledge base			
	Variety	Coherence	Cognitive distance	Convergence
<i>Phases of technological paradigms cycles</i>				
Initial: Discontinuity	Rise	Diminish	Rise	Diminish
Mature: Continuity	Diminish	Rise	Diminish	Rise

Source: Own elaboration

Note: Among technological subclasses that belong to different classes.

3. Representing knowledge bases

The structure of a knowledge base is a network formed by subunits or pieces accumulated sequentially in time (SAVIOTTI, 2009; KRAFFT; QUATRARO; SAVIOTTI, 2011). The pieces of knowledge are the nodes of the network and the combinations between two or more pieces of knowledge are the links. Since the knowledge base allows cumulateness, the transformation of the structure formed by nodes and links represents the evolution of technical progress (SAVIOTTI, 2009; NESTA; SAVIOTTI, 2005, KRAFFT; QUATRARO; SAVIOTTI, 2011).

We can represent knowledge bases using the information contained in patent databases and graph theory. A patent represents a technology in the form of a

new artifact, method or process. Each patent has one or more similar or dissimilar technological domains that represent the interconnected functionalities (sub-technologies). These technological domains are the International Patent Classification (IPC) codes generally linked to areas of scientific-technological knowledge. The higher the level of sub-technology specificity, the higher is the level of disaggregation and the smaller the scope of functionality. To the extent that patents permit to differentiate between technologies and technological domains, patent bases represent knowledge bases that may refer to countries, industries and companies. However, since IPCs are sub-technologies for any level of disaggregation, they are not pieces of knowledge in the theoretical sense, but only a representation. To equalize the pieces of technological knowledge (in the theoretical plane) in IPC codes or observed technological domains (in the empirical plane), two assumptions must be made. The first is that the patent is a technology and the domains are the pieces of knowledge that compose it. The second is that the technology domains contained in a patent share similarities at some level of IPC classification.

Knowledge bases miss some of their attributes when represented with patent data (Table 2). First, patent data are codified knowledge that has an industrial application or functionality. Tacit knowledge is not included and scientific knowledge does not fit in the IPC classification. Second, the technical fields represent domains of knowledge application and not strictly knowledge pieces. Therefore, some attributes of knowledge, such as interdisciplinarity – in terms of similarity or complementarity –, acquire a different meaning. Interdisciplinarity becomes the property of an IPC when it co-occurs with other technical fields that belong to different technological domains, that is, convergence. Third, IPC classifications are only revised when a new technology extends widely. For that reason, IPC classification is very stable over time. As a consequence, new units of knowledge can remain hidden under for a long time and the perception of discontinuities use to be lower than the real (KAY *et al.*, 2014).

In addition to the loss of representativeness of knowledge bases in the theoretical sense, there are also limitations on the use of patent statistics to represent knowledge bases for two reasons. First, because not all innovation efforts result in patents, either because they are non-patentable or poorly protected by patents. Secondly, because patents are also subject to sectoral distribution bias due to the different propensities to patent across industries and technologies.

TABLE 2
Theoretical elements of the knowledge base and its methodological approach

Theoretical concepts	Graph theory	Patents
Base of scientific and technical knowledge (tacit + codified)	NETWORK	Global patent databases Local patent databases (by period, industry, country or technology) Just codified knowledge
Pieces of knowledge (concepts, methods, ideas, experiences)	NODES	<div style="display: flex; align-items: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg); font-weight: bold; padding-right: 5px;">PATENT INFORMATION</div> <div> <p>Technical fields for different levels of aggregation:</p> <p>a) IPC (2-14 digits);</p> <p>b) Others: any aggregation of technology classes according to their functionality</p> <p>Measure: number of occurrences</p> </div> </div>
Links among pieces of knowledge	LINKS	

Source: Own elaboration

In order to handle with the patent data in mathematical terms, we use the graph theory to build the adjacency matrix. The number of technical domains for any level of aggregation gives the dimensions to the matrix (number of rows and columns). We call the number of patents a given IPC i appears in solitary¹ or with the same IPC for a higher disaggregation level in the database ‘co-occurrence $i = j$ ’; and the number of patents that an IPC i occurs together with an IPC j in the database ‘co-occurrence $i \neq j$ ’. Those measures permit to quantify the number of links between IPCs. The adjacency matrix is square, symmetrical and non-directed². The diagonal represents the co-occurrences and the rest of cells represent the frequency of co-occurrences $i \neq j$. P_i indicates the total number of patents in which the IPC i registered at least once as a co-occurrence $i = j$ or $i \neq j$ and P is the total patents in the database. The adjacency matrix is represented as follows:

1 The assignation of one only IPC to a patent is quite rare. We considered them as co-occurrences for the same classification at 4-digit level of aggregation, but it also can be not counted in the diagonal values.

2 Knowledge bases can be also represented using patent citations. In this case, the matrix of interrelations is directed.

Adjacency Matrix

	IPC ₁	IPC ₂	...	IPC _n	Co-occurrences <i>i = j</i>	Co-occurrences <i>i ≠ j</i>	Total Patents
IPC ₁	<i>c</i> ₁₁	<i>c</i> ₁₂	...	<i>c</i> _{1n}	<i>c</i> ₁₁	$\sum_{i=1}^n c_{1n} - c_{11}$	<i>P</i> ₁
IPC ₂	<i>c</i> ₂₁	<i>c</i> ₂₂	...	<i>c</i> _{2n}	<i>c</i> ₂₂	$\sum_{i=1}^n c_{2n} - c_{22}$	<i>P</i> ₂
...
IPC _n	<i>c</i> _{n1}	<i>c</i> _{n2}	...	<i>c</i> _{nn}	<i>c</i> _{nn}	$\sum_{i=1}^n c_{in} - c_{nn}$	<i>P</i> _n
TOTAL	-	-	-	-	-	-	<i>P</i>

As technical progress advances, the adjacency matrix should change in the following sense. First, when new technology domains (IPCs) arise, the number of nodes (lines and columns) increases. Second, when new patents use the same technological domain (IPC), the number of co-occurrences *i = j* increases. Third, as new combinations between IPCs appear, the matrix will be filled outside the diagonal and the co-occurrences will be more diversified. Fourth, the more frequent the interrelationship between two IPC codes, the higher the co-occurrence values in the cells.

4. Database

The database consists of all patent documents available in the European Patent Office’s (EPO) PATSTAT 2015 database and Orbis (BvD) 2017 for 10-years periods³ (1978-87; 1988-1997; 1998-2007; 2008-2016). Even the IPC classification registered modifications period, we used the 2015 revisited classification for the whole period⁴.

The aggregate data show that the 1998-2007 period concentrated 38.7% of the total patents, which reveals that this was the stage of greatest expansion (Table 3). The 2008-2016 period registered 33.7% of the total patents. As long as this is a 9-years period, this stage has to be considered still as provisional information. The

3 The last period has only 9 years because it was impossible to make exactly 10-years periods for the period extent of the available database.

4 The IPC codes which have been deleted in the previous editions/versions do not appear in the current version.

average number of co-occurrences per patent tends to rise from the 1988-1997 period until its maximum value of 0.6 in the 2008-2016. And the average number of co-occurrences $i \neq j$ per patent tends to low from the 1988-1997 until its minimum value of 0.9 in the 2008-2016. The number of domains (nodes) remained between 628 and 634, which indicates that the size of the networks changed little. However, the changes in the average number of links by node was relatively bigger, achieving maximum value in the 1998-2007 period (85). As in this period, the number of patents was also greater, it seems that the evolution of the knowledge base was characterized as a diversified growth.

TABLE 3
Evolution of the knowledge base structure

	1978-1987	1988-1997	1998-2007	2008-2016	Total
N° patents	289,591	669,271	1,347,905	1,172,438	3,479,205
(%/Total)	8.3	19.2	38.7	33.7	100
Number of co-occurrences $i = j$	133,424	287,658	615,302	669,740	-
(Co-occurrences $i = j$) / P	0.5	0.4	0.5	0.6	-
Number of co-occurrences $i \neq j$	456,686	1,315,115	2,285,043	1,092,788	-
(Co-occurrences $i \neq j$) / P	1.6	2.0	1.7	0.9	-
$(co - oc\ i \neq j) / (co - oc\ i = j)$	3.4	4.6	3.7	1.6	-
Number of IPCs (nodes)	628	630	633	634	-
Number of links ⁽¹⁾	31,370	44,476	53,729	42,578	-
Number of links by node	50	71	85	67	-

Source: EPO PATSTAT (2015); Orbis / BvD (2017). Own elaboration.

(1) Only includes links between different IPCs at 4-digits level of aggregation (co-occurrences $i \neq j$)

Gephi software allows the visualization of the knowledge's networks structure (BASTIAN; HEYMANN; JACOMY, 2009). Each period describes very dense networks, that is, with a large number of different domains (nodes) and a vast number of connections (links) (Figures 1, 2, 3 and 4). Other characteristics of the knowledge base are reported according to the size, color, and position of the nodes. The size of each node is proportional to its degree centrality, which is a measure based on the number of links it performs. The larger nodes have a degree above the whole network's average, and the smaller nodes have very few links. The clustering coefficient measures the likelihood of a node to create all the possible links the nodes that are immediate closer.

FIGURE 1
Evolution of the network that represent
the global knowledge base - 1978-1987



FIGURE 2
Evolution of the network that represent
the global knowledge base - 1988-1997



FIGURE 3
Evolution of the network that represent
the global knowledge base - 1998-2007

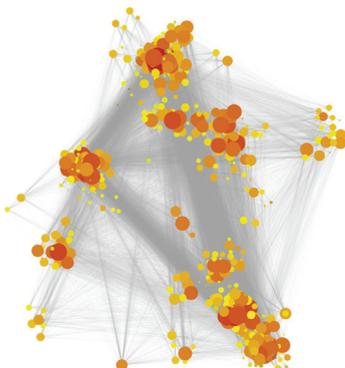
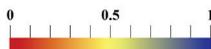
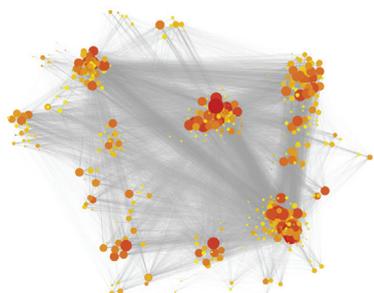


FIGURE 4
Evolution of the network that represent
the global knowledge base - 2008-2016



Source: EPO PATSTAT (2015); Orbis / BvD (2017). Own elaboration.

Note: The color grading indicates the range of the clustering coefficient value for each IPC (nodes). The darker, the closer to 1; the yellow are those closer to 0.5; the red are the closer to 0.

In the knowledge network, when nodes get a coefficient close to one, this indicator points out to the formation of a cluster between certain technological domains and the exclusion of others. The blue color indicates a value equal or very close to one; yellow color indicates values close to 0.5 and red color indicates values equal or close to zero. The blue nodes tend to belong to tightly connected groups, while red nodes tend to form a greater variety of links, proportional to the size of the

network. Few nodes with clustering coefficient close to 1 are present in all periods, but they have the smallest sizes since they tend to create very few links with other nodes. Finally, the node's position in the network depends on how connected they are. Nodes subgroups are created when they concentrate their links among them revealing a higher degree of technological similarity. Alternatively, nodes that are very distant from the network core have fewer links and are more isolated.

As Figure 1, 2, 3 and 4 illustrates, the 1978-1987 period network had many nodes very distant from the core. They do not differentiate by size and red color was predominating, that is, they had similar degree centrality. The 1988-1997 and 1998-2007 bases showed that some technological domains had a distinguished capacity to increase their number of links and were less prone to be restricted to a cluster. Network images confirm that big nodes are more likely to show a red color. Both bases [1988-1997 and 1998-2007] define the emergence and extension of the 80th technological paradigms. The average distance between nodes in the 1978-1987 base is higher than the average distance for the all periods. As the network evolved, nodes connected more and got closer to the network's center. The contraction of the 2008-2016 base can not be attributed to a smaller number of nodes, but to the distribution of the links between them. After the micro-paradigms of KETs emerged, the number of links grew again, but following the 80s paradigms maturity processes. That means that their entrance consolidated the preexisting technological trajectories that formed groups and made the variability in the size of nodes less distinguishable.

5. Analysis of variety, coherence and cognitive distance

The adjacency matrix allows elaborating the indicators for variety, coherence and cognitive distance. Informational entropy index measures variety as the concentration of co-occurrences $i \neq j$ by technological domain at any specific level of aggregation (KRAFFT; QUATRARO; SAVIOTTI, 2011; QUATRARO, 2010; FRENKEN; VAN OORT; VERBURG, 2007; FRENKEN; NUVOLARI, 2004). This indicator is formalized as follows:

$$TV \equiv (X, Y) = \sum_i \sum_j p_{ij} \log_2 \left(\frac{1}{p_{ij}} \right)$$

Where $p_{ij} = c_{ij}/P$ is the probability of co-occurrence of the pair of IPCs $i - j$, calculated as the ratio between the number of $i - j$ co-occurrences and the

total patents of the period. The total variety will be greater the lower the probability of co-occurrence of a specific pair. The total variety index can be broken down into related and unrelated variety. Unrelated variety (UV) measures the concentration of the co-occurrences $i \neq j$ between unrelated domains of knowledge at lower levels of aggregation. This kind of variety can be associated with disruptive technologies and evolution by discontinuity. Related variety (RV) measures the concentration of the co-occurrences $i \neq j$ between related domains at higher levels of aggregation, that is, that belong to the same technological class. If the i and j domains belong to different classes [g and z], such that $i \in S_g$ and $j \in S_z$, then the probability of co-occurrence $i \neq j$ is:

$$p_{gz} = \sum_{i \in S_g} \sum_{j \in S_z} p_{ij}$$

While the probability of subsets co-occurrence is:

$$H_{gz} = \sum_{i \in S_g} \sum_{j \in S_z} \frac{p_{ij}}{p_{gz}} \log_2 \left(\frac{1}{p_{ij}/p_{gz}} \right)$$

Therefore:

$$RV = \sum_{g=1}^G \sum_{z=1}^Z p_{gz} H_{gz}$$

$$UV \equiv H_Q = \sum_{g=1}^G \sum_{z=1}^Z p_{gz} \log_2 \frac{1}{p_{gz}}$$

$$TV = H_Q + \sum_{g=1}^G \sum_{z=1}^Z p_{gz} H_{gz}$$

Coherence measures complement the analysis of variety indicators showing how the network structure combines knowledge. There are three indicators for coherence: *density*, *average clustering coefficient* and the *average degree* (WATTS; STROGATZ, 1998; BARABÁSI; ALBERT, 1999; JACKSON, 2010). Density (D) is the rate

between the number of existing links and the total possible co-occurrences of the matrix as follows:

$$D = \frac{2l^*}{n \times (n - 1)}$$

Where l^* counts the number of pairs of co-occurrences (c_{ij}) in the adjacency matrix; n is the number of technological domains (rows = columns). The maximum value of network density is equal to 1 when the matrix is filled by all the possible pairs of co-occurrences. A density value close to zero reveals relatively few connections between the technological domains.

In addition to directed links (c_{ij}), the coherence of the network is also determined by indirect links that lead to the formation of groups among technological classes. When the nodes $j \neq i \neq k$ combine, they become neighbors and create a complete sub-network filling the cells $c_{ij}, c_{ik} \in c_{jk}$, that is, they are connected by all the possible links among them. When this happens, they are more similar between them than with other nodes they are not linked directly with. The average clustering coefficient captures in which extent a node is connected to subnetworks with at least two nodes within them (WATTS; STROGATZ, 1998). The clustering coefficient C_i for a i -node is the ratio between the number of co-occurrences $i \neq j$ formed by i within its K neighbors (N) and the maximum possible number of co-occurrences that it can create with its K neighbors $K * (K - 1)/2$ as follows:

$$C_i = \frac{N_i}{K_i * (K_i - 1)/2}$$

The average clustering coefficient of a network is an average of the individual clustering coefficients C_i of all the n -nodes (IPCs) in the database:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

The average clustering coefficient is equal to 1 if all the IPCs $i \neq j$ co-occur, this is, when the adjacency matrix is fully filled; and it is equal to 0 if there are no IPCs $i \neq j$ co-occurring.

Degree centrality is a count of the total number of co-occurrences $i \neq j$ for each domain in the network. The average degree of an undirected graph is

measured by the division of the sum of all nodes' degree by the total number of nodes. When the technological domains create new co-occurrences $i \neq j$, they fill the co-occurrence matrix and raise the average degree and so the complexity of the knowledge base.

The average path length and the diameter measure the cognitive distance. Consider the co-occurrence $i \neq j$. The shortest possible path length between $i - j$ is equal to 1. Consider now two domains not directly linked to each other but related with a third domain $[k]$, such as $[i - j; j - k]$. The path length between i and k is the sum of the direct links between the domains that intermediate their connection. For example, if $j \neq i \neq k$ and there are one combination $i - j$ and one combination $i - k$, the path $i - k$, $w(j, k) = 2$, which is the sum of the distances $w(i, j) = 1$, and $w(i, k) = 1$. In aggregate terms, the average path length indicates the average distance between all the different $i - j$ pairs in the adjacency matrix. This indicator considers whether the connections are making easier the intermediation between different technologies. Finally, the diameter (d) is defined as the largest distance between two $i - j$ domains, such that $d = \max w(i, j)$, considering the distances between all pairs of base IPCs. A lower diameter of the knowledge network indicates that the dissimilarity between technological domains got lower.

Table 4 shows the results of the selected indicators. Between 1978-1987 and 1988-1997 periods, coinciding with the emergence and extension of the 80s technological paradigms, there was an increase in the number of IPCs, which lead to a significant increase in the variety of knowledge. Related variety grew more than unrelated variety between 1978-1987 and 1988-1997, that is, diversification occurred more intensely in IPCs that belong to the same areas of knowledge (for example, inside Pharmaceuticals or Telecommunications). This result reveals that the knowledge base diversifies through incremental innovations. In the transition from 1988-1997 to 1998-2007, the number of patents almost doubled and there was also an increase in the number of nodes (Table 3). Variety rate decreased in the 1998-2007 period as result of lower unrelated and related variety showing that even though the network got bigger there was a relatively smaller impulse for the formation of new links. Therefore, the 1998-2007 knowledge base links are more concentrated and largely repeat the previous structure; the knowledge base evolved more by cumulativity than by discontinuity. Between 1998-2007 and 2008-2016 periods, there were significant declines in the variety, both related and unrelated.

TABLE 4
Knowledge network properties

	1978-1987	1988-1997	1998-2007	2008-2016	
Variety	Unrelated variety (UV)	10.2	11.1	10.5	8.4
	Related variety (RV)	8.1	9.0	7.9	5.8
	Variety rate (RV/UV)	0.8	0.8	0.7	0.7
Coherence⁽¹⁾	Density	0.16	0.22	0.27	0.21
	Average clustering coefficient	0.52	0.58	0.62	0.59
	Average degree	100	141	170	134
Cognitive distance	Average path length	1.90	1.80	1.74	1.82
	Diameter	4	4	3	4

Source: EPO PATSTAT (2015); Orbis / BvD (2017). Own elaboration.

(1) Values vary between 0 and 1

Regarding coherence, the density indicator was very low in the initial period (0.16) for a maximum value of 1. This value increased up to 0.22 between the first two periods and then up to 0.27 in the 1998-2007 period. In the last period [2008-2016], it fell to 0.21. All the technological domains presented at least one relation and the average of relations per node got higher until the 1998-2007 period (170). This evolution confirms the formation of new links between pairs of IPCs. The knowledge base experienced an integration process that seems to have slowed down in the 2008-2016 period coinciding with the emergence of the KETS. Nevertheless, this observation has to be still confirmed when the latter period completes 10 years. The average clustering coefficient, which was already high in the 1978-1987 period (0.52), increased more until the 1998-2007 period (0.62). These measures reveal a tendency to increase cohesion between all the IPCs, that is, the knowledge base evolves more cumulatively with more local convergence. In other words, the network evolved concentrated in specific clusters.

The cognitive distance or dissimilarity measured by the average path length started out being relatively low (1.90) in the 1978-1987 and fell further over time from 1.80 to 1.74 in the 1998-2007. This result is closely related to the appearance of new links until 1998-2007, which made the cognitive distance shorter and the

knowledge more interrelated in the network. The increase in the number of new links, without a significant increase in the number of new nodes, reduced the overall dissimilarity until 1998-2007. This was because some technology domains (nodes) increased the variety of their connections. For a medium-sized network of 630 nodes, the diameter values were relatively small, which reflects the large interconnection between the technological domains of the network. As a result, technological dissimilarity was reduced in 1998-2007.

In summary, in the passage from the 1980s to the 1990s, with the emergence of the 1980s technological paradigms, the knowledge base increased in variety and coherence simultaneously and cognitive distance reduced. Afterwards, the base tended to specialize as an accumulative effect. Since the 2000s, the network becomes more integrated, as showed by coherence measures. The emergence of the new paradigms (KETs) leads to the appearance of new nodes and links, but with a fall in the variety. That observation indicates that the connections between domains became more concentrated in specific technology fields. The results also indicate that this tendency to concentration stands during the 2008-2016 period. The drop in the co-occurrences $i \neq j$ per domain explains lower variety and coherence in the knowledge structure. Although that is only a provisory outcome, it is not expected that the lower number of patents be an explanation for this observation.

6. Analysis of convergence

When two or more dissimilar pieces of knowledge combine to create a technology, we say they converge. In empirical terms, technological convergence happens when dissimilar IPCs (pieces of knowledge) co-occur in the same patent (technology).

The analysis of convergence has the purpose to identify in the base of knowledge how the 80s technological paradigms and the KETs evolved combining with dissimilar technological fields to create new technologies. To do that, we first create for each period the subnetworks relative to each 80s paradigm (biotechnology, new materials, microelectronics, informatics and computers, telecommunications, semiconductors and audio-visual technology) and KETs micro-paradigms (advanced manufacturing, advanced materials, micro and nanoelectronics, industrial biotechnology, photonics and nanotechnology) (VAN DE VELDE *et al.*, 2012).

We identify the patents corresponding to each of the 80s paradigms using the aggregation of IPCs proposed by the WIPO technology concordance table in thirty-five fields of technology (WIPO Statistics Database, 2019). Patents activities in

KETs are identified based on a list of IPC codes that cover new technologies directly representing one of the six KETs (VAN DE VELDE *et al.*, 2012; ASCHHOFF *et al.*, 2010)⁵.

Once created the subnetworks, we use the related and unrelated variety indicators to measure convergence. Related variety (RV) measures how the knowledge subnetwork of each micro-paradigm diversified along time considering all the technological fields, whether they are dissimilar or not. Unrelated variety (UV) measures also the evolution of the subnetwork diversification, but only towards dissimilar technical fields, that is, combination with IPCs that belongs to another field of technology, given the aggregation of technology used.

Table 5 compares the results for the 80s paradigms, and shows that UV prevails in relation to RV, indicating a predominant expansion and diversification outside the paradigms, that is, knowledge base grew with *outwards convergence*. RV and UV grew from 1978-1987 to 1988-1997, except for semiconductors that remained relatively stable, which confirms a rise of convergence inside and among the paradigms. For both periods, the highest value for RV and UV belonged to new materials, which encompasses all basic materials chemistry technologies. In the transition from 1988-1997 to 1998-2007, the number of patents grew, but mostly the number of links, as seen previously (see Table 3). In the 1998-2007 base, audio-visual technology and new materials showed a fall of RV and UV in relation to the previous period, indicating a concentration of their links. Biotechnology, informatics and computers, semiconductors and telecommunications continued to increase UV expanding its new connections with technologies outside their paradigms. The 2008-2016 base shows a reduction in variety in all paradigms, with the exception of microelectronic's RV. However, the decline is more prominent in RV than in UV. That means that even with a reduction of the number of links, the links between the knowledge belonging the 80s paradigms and other dissimilar pieces of knowledge were less volatile, indicating the persistence of the preexisting technological trajectories. In the last-period base, the reduction of the total number of patents does not justify a reduction of the links. The strength of a certain *inwards convergence* is due to: (i) the increase of the number of co-occurrences $i = j$; (ii) the higher frequency of the same co-ocurrences; (iii) the disconnection between domains that grew considerably (for example, in the 2008-2016 period, certain links between telecommunications domains and chemistry domains ceased to exist completely).

5 One should note that there is some overlap between KETs. Most importantly, some IPC codes in the field of nanotechnology are also assigned to micro-/nanoelectronics and new materials. There is also a minor overlap between photonics and micro-/nanoelectronics.

TABLE 5
Convergence analysis for the 1980s' technological paradigms

1980s Paradigms (%) ⁽¹⁾	1978-1987		1988-1997		1998-2007		2008-2016	
	RV	UV	RV	UV	RV	UV	RV	UV
<i>Audio-visual technology</i>	0.33	0.62	0.44 (0.33)	0.77 (0.24)	0.41 (-0.07)	0.71 (-0.08)	0.20 (-0.51)	0.43 (-0.39)
<i>New materials</i>	0.75	0.84	0.86 (0.15)	0.90 (0.07)	0.58 (-0.33)	0.69 (-0.23)	0.36 (-0.38)	0.51 (-0.26)
<i>Biotechnology</i>	0.34	0.43	0.59 (0.74)	0.63 (0.47)	0.53 (-0.10)	0.66 (0.05)	0.21 (-0.60)	0.38 (-0.42)
<i>Informatics and computers</i>	0.21	0.43	0.35 (0.67)	0.61 (0.42)	0.41 (0.17)	0.77 (0.26)	0.24 (-0.41)	0.57 (-0.26)
<i>Microelectronics</i>	0.67	0.75	0.70 (0.04)	0.76 (0.01)	0.70 (0.00)	0.73 (-0.04)	0.70 (0.00)	0.70 (-0.04)
<i>Semiconductors</i>	0.08	0.36	0.11 (0.38)	0.44 (0.22)	0.11 (0.00)	0.45 (0.02)	0.07 (-0.36)	0.34 (-0.24)
<i>Telecommunications</i>	0.28	0.56	0.45 (0.61)	0.80 (0.43)	0.60 (0.33)	1.08 (0.35)	0.32 (-0.47)	0.78 (-0.28)

Source: EPO PATSTAT (2015); Orbis / BvD (2017). Own elaboration.

(1) Rate of growth between consecutive periods.

In relation to the 2000s micro-paradigms (the KETs), the convergence analysis displayed in Table 6 reveals some differences in relation to the 80s paradigms results. Previously to the first decade of 2000, the 1978-1987 base show that the nodes belonging to the KETs technologies developed *inwards*, that is, inside their own subset of related technologies. RV was higher than UV, particularly in advanced manufacturing. Advanced manufacturing technologies stand out from the other KETS since its knowledge base characterizes by a wider scope of combinations.

All the KETs' knowledge bases increased the RV in the 1988-1997 period, except nanotechnology. The UV also increased, especially for the industrial biotechnology⁶, which indicates some *outwards convergence*. In the next period (1998-2007), some KETs bases of knowledge showed a reduction in total variety, but the drop in RV was stronger than UV, which points out that *outwards convergence* persisted more. Micro and nanoelectronics, photonic and nanotechnology reached higher values of UV than RV, which reflects paths of technological trajectories associated with other

6 There is some overlap between this technological classification and the one used for Biotechnology in the 80s paradigms classification, but the most important difference is that some of the KETs classification codes are IPC 7 digits, therefore, the information extracted has a higher level of technological specificity.

technologies. Photonics and nanotechnology increased also RV, while industrial biotechnology only increased through UV. Advanced materials reducing both, RV and UV from 1998-2007 period, that is, it concentrated their connections links reflecting a no-convergence path. Photonics and nanotechnology registered the lowest variety, but both tended sensitively to increase their scope of combinations during all the periods, reducing considerably their distance with other technologies in the 2008-2016 period. Photonic was the only one that continued growing in both, RV and UV, until the last period, revealing a path of *inwards* and *outwards convergence*.

TABLE 6
Convergence analysis for the 2000s' technological paradigms

2000s Micro-paradigms (KETs) (%) ⁽¹⁾	1978-1987		1988-1997		1998-2007		2008-2016	
	RV	UV	RV	UV	RV	UV	RV	UV
<i>Advanced Manufacturing</i>	2.88	0.71	2.91 (0.01)	0.70 (-0.01)	2.19 (-0.25)	0.61 (-0.13)	1.57 (-0.28)	0.54 (-0.11)
<i>Advanced Materials</i>	0.54	0.33	0.67 (0.24)	0.34 (0.03)	0.48 (-0.28)	0.26 (-0.24)	0.29 (-0.40)	0.21 (-0.19)
<i>Micro and nanoelectronics</i>	0.20	0.25	0.25 (0.25)	0.28 (0.12)	0.24 (-0.04)	0.27 (-0.04)	0.14 (-0.42)	0.23 (-0.15)
<i>Industrial Biotechnology</i>	0.34	0.22	0.63 (0.85)	0.32 (0.45)	0.63 (0.00)	0.34 (0.06)	0.22 (-0.65)	0.21 (-0.38)
<i>Photonic</i>	0.04	0.07	0.06 (0.50)	0.08 (0.14)	0.07 (0.17)	0.08 (0.00)	0.08 (0.14)	0.11 (0.38)
<i>Nanotechnology</i>	0.00	0.00	0.01 -	0.01 -	0.03 (2.00)	0.03 (2.00)	0.02 (-0.33)	0.03 (0.00)

Source: EPO PATSTAT (2015); Orbis / BvD (2017). Own elaboration.

(1) Rate of growth between consecutive periods.

In sum, the KETs's knowledge bases departed from a low degree of diversification associated to an incipient stage characterized by the search of applications of a new body-of-knowledge. Afterwards, specific trajectories defined a still more restricted set of domains towards KETs technology domains established and concentrated their links. Nevertheless, the most recent period reveals a new phase of KETs technologies characterized by an increasing variety by convergence (unrelated), especially in photonics and nanotechnology. As this is a recent path, the tendency to diversification by convergence is not still sufficient to alter the total variety of the whole knowledge base.

7. Conclusions

The aim of this work was to analyze the evolution of the knowledge base according to different phases of the evolution cycle of the seven technological paradigms in the 1980s and the Key Enabling Technologies (KETs). The complexity that characterizes knowledge bases was studied from its four interconnected properties: variety, coherence, cognitive distance and convergence. The article assumes by hypothesis that at the initial stage of new paradigms, the knowledge base is characterized by an increase in variety, a reduction in coherence and, as a result, a reduction in the density of the network in general. To do that, we used patent statistics and graph theory to build a network that represent the bases of knowledge in different 10-years periods from 1978 to 2016. The network we made is, in this sense, just a representation of a technology base knowledge that considers as nodes the technological domains corresponding to 4-digits level of aggregation of the IPC Classification.

The results confirm the hypotheses only partially. The initial period characterized by the emergency of new paradigms coexisting with the maturity of the old ones. There was an increase in coherence due to the increase of the connections that widely offset the effect of the emergence of new domains. From the 1980s on, the maturation process advances. The transition from 1988-1997 to 1998-2007 showed a decrease in variety and in cognitive distance, an increase in coherence, which was expected considering the path of knowledge accumulation. With the fall of the variety, the network did not integrate in a balanced way, which was revealed in the growth of interconnectivity detected by the agglomeration coefficient. The maturation was characterized by the emergence of new relationships between IPCs within the same area of knowledge (by increase of related variety or by cumulativity), but also, although with less importance, between different areas (by unrelated variety or discontinuity). Considering the cumulative and path-dependent character of knowledge, these observations reveal that there was continuity in the evolution the technological paradigms, but also the disuse of some technological trajectories, as certain links have lost their relative importance.

The new paradigms demonstrated a tendency to increase interconnectivity, which points out towards convergence as the main characteristic of the maturation stage of their technological cycles. The 80s paradigms showed a predominant expansion and diversification outside the paradigms, that is, the knowledge base grew with outwards convergence. Even in the 2008-2016 base when the total variety contracted for all the paradigms, the links between the knowledge belonging to the

80s paradigms and other dissimilar pieces of knowledge were less volatile, indicating persistence of technological trajectories. The KETs's knowledge bases departed from a low degree of diversification associated to an incipient stage characterized by the search of applications of a new body-of-knowledge into specific trajectories increasing variety by convergence (unrelated). At the most recent period most KETs technologies also tended to concentrate their links.

The outward path of convergence reveals also the existence of tradeoffs between technological domains that can involve constraints to technical progress. For example, the application of photonics and nanotechnology in the development of other technologies can mean a strong dependence on the performance of these other technological trajectories to their own maturation process. Innovation policies should pay attention to this coevolution and expect tradeoffs and spillovers among technologies answering this kind of questions: what kind of technologies are in the technological frontier? What kind of complementary and dissimilar technological domains have to be supported to guarantee the advance of the technological frontier?

Finally, it is necessary to stand up two assessments. Firstly, the results for the 2008-2016 period in comparison with the others must be considered still provisional, given that the period has not completed 10 years yet. Nevertheless, even the data are not still available to complete the decade; it seems that the general observed tendency will remain unchanged. The contraction of the base of knowledge in the last period reveals the disappearance of weak or eventual connections and the concentration on the stronger connections, following the eighties-paradigms' trajectories. At the same time, and from the beginning of the 2000s, new micro-paradigms emerged: the KETs. This new wave of disruptive technologies also can have played a role towards the contraction of the knowledge base. Until very recently, the KETs' domains mainly trended to connect with their similar, which drove more to an endogamic concentration than to an exogamic diversification.

Secondly, the structure of the knowledge base depends on the level of aggregation of the IPC classification. The higher the level of aggregation, the greater the number of nodes and the relationships between them, which increases the complexity and stability of the knowledge base. In this sense, other specificities can emerge carried out by a more complex structure that cannot be observed at 4-digit level of aggregation.

References

- ARTHUR, W.B. *The nature of technology: What it is and how it evolves*. New York: Simon & Schuster, 2009.
- ASCHHOFF B.; CRASS, D.; CREMERS, K.; GRIMPE, C.; RAMMER, C.; BRANDES, F.; MONTALVO, C. *European competitiveness in key enabling technologies*. Mannheim, Germany, Centre for European Economic Research (ZEW), 2010.
- BARABÁSI, A.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509-512, 1999.
- BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. *Icwm*, v. 8, n. 2009, p. 361-362, 2009.
- DOSI, G. *Technical change and industrial transformation: the theory and an application to the semiconductor industry*. UK: Palgrave Macmillan, 1984.
- EPO PATSTAT. EPO Worldwide Patent Statistical Database. European Patent Office (EPO) 2015 Spring Edition. Version 5.03.
- FLEMING, L.; SORENSON, O. Technology as a complex adaptive system: evidence from patent data. *Research Policy*, v. 30, n. 7, p. 1019-1039, 2001.
- FREEMAN, C.; PEREZ, C. Structural Crises of Adjustment, Business Cycles and Investment Behavior. In: DOSI, G. et al. (ed.). *Technical Change and Economic Theory*. London: Pinter, 1988. p. 39-62.
- FRENKEN, K.; NUVOLARI, A. Entropy statistics as a framework to analyse technological evolution. In: FOSTER, J.; HÖLZL, W. (ed.). *Applied evolutionary economics and complex systems*. Cheltenham: Edward Elgar Publishing, 2004. p. 95-133.
- FRENKEN, K.; VAN OORT, F.; VERBURG, T. Related variety, unrelated variety and regional economic growth. *Regional studies*, v. 41, n. 5, p. 685-697, 2007.
- HUENTELER, J.; SCHMIDT, T.S.; OSSENBRINK, J.; HOFFMANN, V.H. Technology life-cycles in the energy sector – Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, v. 104, p. 102-121, 2016.
- JACKSON, M.O. *Social and economic networks*. Princeton University Press, 2010.
- JOO, S.; KIM, Y. Measuring relatedness between technological fields. *Scientometrics*, v. 83, n. 2, p. 435-454, 2010.

KAY, L.; NEWMAN, N.; YOUTIE, J.; PORTER, A. L.; RAFOLS, I. Patent overlay mapping: Visualizing technological distance. *Journal of the Association for Information Science and Technology*, v. 65, n. 12, p. 2432-2443, 2014.

KIM, E.; CHO, Y.; KIM, W. Dynamic patterns of technological convergence in printed electronics technologies: patent citation network. *Scientometrics*, v. 98, n. 2, p. 975-998, 2014.

KRAFFT, J.; QUATRARO, F.; SAVIOTTI, P.P. The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology*, v. 20, n. 5, p. 445-475, 2011.

NESTA, L.; SAVIOTTI, P. P. Coherence of the knowledge base and the firm's innovative performance: evidence from the US pharmaceutical industry. *The Journal of Industrial Economics*, v. 53, n. 1, p. 123-142, 2005.

ORBIS. Orbis global database. Bureau van Dijk Electronic Publishing (BvD). Accessed in: November 2017.

QUATRARO, F. Knowledge Coherence, Variety and Productivity Growth: Manufacturing Evidence from Italian Regions. *Research Policy*, v. 39, n. 10, p. 1289-1302, 2010.

SAHAL, D. Alternative conceptions of technology. *Research Policy*, v. 10, n.1, p. 2-24, 1981.

SAHAL, D. Technological guideposts and innovation avenues. *Research Policy*, v. 14, n. 2, p. 61-82, 1985.

SAVIOTTI, P.P. Knowledge networks: structure and dynamics. In: PYKA, A.; SCHARNHORST, A. (ed.). *Innovation networks: new approaches in modelling and analyzing*. Berlin Heidelberg: Springer-Verlag, 2009. p. 19-41.

VAN DE VELDE, E.; RAMMER, C.; GEHRKE, B., DEBERGH, P., SCHLISSLER, P.; WASSMANN, P. *Key Enabling Technologies (KETs) Observatory*. Second Report. Brussels, EC DG GROW, 2015.

VERSPAGEN, B. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, v. 10, n. 1, p. 93-115, 2007.

WALTMAN, L.; VAN ECK, N.J.; NOYONS, E.C. A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, v. 4, n. 4, p. 629-635, 2010.

WATTS, D.J.; STROGATZ, S.H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440, 1998.

WIPO. *Guide to the International Patent Classification*. Version 2019.

WIPO. *WIPO Economics & Statistics Related Resources 8*. World Intellectual Property Organization - Economics and Statistics Division, Last update: March 2018.

YAN, B.; LUO, J. Measuring technological distance for patent mapping. *Journal of the Association for Information Science and Technology*, v. 68, n. 2, p. 423-437, 2017.

Contribution of each author:

A. Theoretical and conceptual foundations and problematisation: Ana Urraca Ruiz

B. Data research and statistical analysis: Vanessa de Lima Avanci

C. Elaboration of figures and tables: Vanessa de Lima Avanci

D. Drafting and writing of the text: Ana Urraca Ruiz and Vanessa de Lima Avanci

E. Selection of bibliographical references: Ana Urraca Ruiz and Vanessa de Lima Avanci

Conflict of interest: the authors declare that there is no conflict of interest.

Source of funding: the authors declare that there is no source of funding.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License CC-BY, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.