department of economics
"S. Cognetti de Martiis"

Via Po, 53 – 10124 Torino (Italy)
Tel. (+39) 011 6704917 - Fax (+39) 011 6703895
URL: http//www.de.unito.it

WORKING PAPER SERIES

EVOLUTION OF THE KNOWLEDGE BASE IN KNOWLEDGE INTENSIVE SECTORS

Jackie Krafft, Francesco Quatraro, Paolo Saviotti

Dipartimento di Economia "S. Cognetti de Martiis"

LEI & BRICK - Laboratorio di economia dell'innovazione "Franco Momigliano"
Bureau of Research in Innovation, Complexity and Knowledge, Collegio Carlo Alberto

Working paper No. 06/2009

Università di Torino

Evolution of the Knowledge Base in Knowledge Intensive Sectors

Jackie Krafft*, Francesco Quatraro**, Paolo Saviotti***

* University of Nice Sophia Antipolis, CNRS GREDEG
** University of Nice Sophia Antipolis, CNRS GREDEG, and BRICK, Collegio Carlo Alberto
*** INRA GAEL, and University of Nice Sophia Antipolis, CNRS GREDEG

ABSTRACT. The paper develops an evolutionary framework to investigate and compare the dynamics of knowledge base across three knowledge intensive sectors, i.e. biotechnology, telecommunications and electronics. Knowledge is understood as collective good featured by a co-relational and a retrieval-interpretative structure. The internal structure of knowledge is described as a network the nodes of which are small units within traces of knowledge like patent documents, connected by links determined by their joint utilisation. We thus derived a number of properties like variety, coherence and cognitive distance, by using co-occurrence matrixes referring to the citation of technological classes within patent documents. Empirical results show the existence of interesting and meaningful differences across the sectors, which are linked to the different phases of lifecycles the industries have undergone in the period of observation.

JEL Classification Codes: O33

Keywords: Knowledge Base, Variety, Coherence, Industry lifecycles

1 Introduction[1]

The economic systems of advanced capitalistic societies have been facing a gradual process of transition towards the so-called knowledge-based economy. In this context the creation and utilisation of knowledge become the key factors affecting the competitiveness of firms, regions and countries (Freeman and Soete, 1997).

In view of this, the study of the mechanisms of knowledge production has received renewed attention in the last decade, and a considerable effort today is dedicated to characterise the knowledge base of different sectors in the economy and to detect its impact on firm performance and on industrial organization (Breschi, Lissoni, and Malerba, 2003; Nesta and Saviotti, 2005; Krafft, 2006; Corrocher et al., 2007; Antonelli, 2008). These studies provide useful evidence of how much pervasive the production of knowledge is in shaping the economic performances of firms within a specific sector (Piscitello, 2004; Nesta, 2008). Yet, only a few efforts are dedicated to investigate the specific evolution of knowledge bases in knowledge-intensive sectors (or KISs) and there are no empirical contributions adopting a sectoral comparative approach to the issue (Grebel, Krafft, Saviotti, 2006).

This paper investigates and compares the dynamics of the knowledge bases of three key knowledge-intensive industries, i.e. biotechnology, telecommunications and electronics. Our objective is to establish the mechanisms by means of which knowledge is created and used in these knowledge intensive sectors, and characterize their pattern of evolution over time. In particular a number of properties are identified, namely (i) variety (related and unrelated), (ii) coherence, and (iii) cognitive distance, which can describe the evolution of the internal structure of the knowledge base in each sector. The analysis is carried out by adopting an evolutionary viewpoint, in which knowledge is understood as a collective good.

According to evolutionary theory, technological knowledge and innovation are key factors triggering economic development. The introduction of technological innovations brings about new variety in the economic system, providing the bases for restless economic growth (Metcalfe, 2002). New knowledge stems from research efforts, start by exploring new regions of knowledge space, thus constructing the basic conceptual infrastructures which can later lead to the exploitation of the knowledge thus created. Exploitation dynamics occurs within current technological trajectories, as long as they prove to be profitable. Although new trajectories are often initiated by radical innovations subsequent innovations are mainly incremental, and recombination is more likely to involve related technological fields. However, technological opportunities tend to exhaust over time, and market for products based on them are likely to get saturated as well. Radically new technologies then emerge out of a process of random screening across a wide body of technological domains, in search for new profitable recombinations. This phase is characterized by the recombination of loosely

[1] This work is part of a research project funded by Agence Nationale de la Recherche (contract number: ANR JCJC06_141306, “Knowledge Intensive Sectors: Models and Evidence”). Preliminary versions have been discussed at the EAEPE conference “Labour, Institutions and Growth in a Global Knowledge Economy” held in Rome on 6-8 November 2008, the workshop “The Dynamics of Knowledge and Innovation in Knowledge-Intensive Industries” oganized by the BRICK and the GREDEG at the Collegio Carlo Alberto on 18-19 December 2008, and at the workshop “The role of inventors and patents: Analysis and methodological issues”, organized at the LEFI-ESDES in Lyon on 29th January 2009. The authors wish to thank Cristiano Antonelli, Ron Boschma, Aldo Geuna and Marco Valente for their comments, as well as the funding of Collegio Carlo Alberto through the BRICK research centre, as well as the support of CNRS through the GREDEG research centre.

related technologies, and marks a discontinuity in the evolution of the knowledge base (Saviotti, 1996). A discontinuity occurs when completely new concepts and theories are introduced into a given field of science or an industrial activity. Such new concepts may be borrowed from other disciplines or be created in the exploration of completely new regions of knowledge space. We expect discontinuities to exert a very powerful influence on the internal dynamics of knowledge and on its possible industrial applications.

The collective nature of knowledge refers to its intrinsic cumulativeness and to the processes of recombination underlying its production. This allows us to understand knowledge as a retrieval-interpretative and a co-relational structure. In view of this, the properties characterizing the internal structure of knowledge bases, and their evolution, can be derived by using co-occurrence matrixes referring to the citation of technological classes within patent documents (Saviotti, 2004 and 2007; Grebel, Krafft, Saviotti, 2006).

In this paper we use patent data to study the evolution of the KBs of biotechnology, telecommunications and electronics. According to OECD STI scoreboard (OECD, 2007), such sectors may be defined as “high technology and knowledge intensive sectors”, along with aerospace. Such classification draws upon a number of different indicators, like R&D intensity, the share of human capital employed in science and technology based activities, patent intensity, technology trade, and so on and so forth. The data we use in this paper indeed show that their rate of patent production is clearly higher than that of the average industrial sector. We may consider this evidence sufficient for our purposes and proceed to study how our sectors create and use knowledge.

Empirical results show the existence of interesting and meaningful differences across the biotechnology, telecommunications and electronics sectors. This holds both with reference to the relative levels of related and unrelated variety, knowledge coherence and cognitive distance, and with respect to the evolution of the same indexes over time. Such differences are linked to the different phases of lifecycles the industries have undergone in the period of observation. In sum, this paper makes three contributions to the existing literature. First, it defines three properties of the knowledge base in knowledge intensive sectors. Second, it elaborates a link with the phases of exploration and exploitation that characterize an industry life cycle. Third, it provides a distinction between related and unrelated variety, which can be used to characterize whether KISs are still in an initial stage of development or progressively moving to a maturity stage.

The rest of the paper is organized as follows. Section 2 outlines the theoretical framework underlying this study, and provides three working hypotheses. Section 3 describes the methodology we used in our study, and provides definitions of the variables that proxy the internal structure of knowledge base. Section 4 presents the data. In Section 5, we show our empirical evidence. Empirical evidence exhibits some commonalities in the three sectors. In each sector, it is possible to distinguish between the phase of random screening (exploration) and organized search (exploitation) by looking at the evolution of the distribution of co-occurrences across technological classes. But empirical evidence also captures differences in the evolution of the knowledge bases of the different sectors that we interpret. Section 6 discusses the results and concludes.

2 Theoretical Framework

In spite of the recognized importance of knowledge as a potential determinant of growth, economics has not yet developed an adequate representation of knowledge comparable to that of physical capital. Past attempts to measure and to model knowledge have included the use of patents and publications either as indicators of knowledge production or as inputs to a knowledge production function (Griliches, 1979 and 1990; Narin, 1994). While these attempts were important and they contributed to improve our understanding of the economics of knowledge, they relied on traces and not on actual measurements of knowledge itself. By traces of knowledge we mean here phenomena that we know are related to knowledge but according to mechanisms we do not fully understand. To proceed beyond this stage we need adequate conceptual definition and representation of knowledge itself. In the past such a task has been attempted by philosophers and epistemologists. The emergence of knowledge based society changes the boundaries of economic phenomena and forces economics to include fundamental considerations about knowledge. The representation of knowledge which is required in economics must allow us to treat in a comparative way the various types of knowledge which are created and used in different institutions, ranging from public research organizations to private firms. By establishing a continuity amongst these different types of knowledge economics would then pursue the opposite approach with respect to epistemologists attempting to find the demarcation between science and other forms of knowledge.

The representation we require is not necessarily a complete one but it can be based on a number of properties of knowledge. Examples of these properties are the following (Saviotti, 2004, 2007):

(i) Knowledge is a co-relational structure
(ii) Knowledge is a retrieval or interpretative structure

According to (i) knowledge establishes co-relations or connections between variables or concepts. According to (ii) knowledge allows us to recover types of knowledge similar to those we already knew endowing us with an absorptive capacity for them (Cohen and Levinthal, 1990). From these two properties we can deduce that knowledge can be represented as a network the nodes of which are variables, connected by links determined by the joint utilisation of different variables. We can also expect that the overall network of knowledge will never be fully connected since new variables are likely to be created in different regions of knowledge space, corresponding to different disciplines, before all the possible connections are established. In other words, the rates of creation of new nodes in the network of knowledge cannot be expected to coincide at all times with the rate of creation of links. As a consequence network density becomes a relevant variable to characterize the dynamics of knowledge.

The possibility to represent knowledge as a network provides an adequate conceptual foundation for the study of processes of knowledge generation and utilization in firms and industries. To identify all the variables and the connections present in the knowledge base of a firm would be a prohibitively expensive task. An approximate version can then consist of identifying relatively 'small' units of knowledge and their connections. We identify these 'small' units within the traces of knowledge which have been used so far, such as patents and publications. However, in principle we can use as source of information any text describing the type of knowledge which we intend to study. Possible 'small' units of knowledge are (i) the technological classes attached to each patent and (ii) the themes which can be identified in texts by means of linguistic engineering procedures. Technological classes are more easily available and simpler to use. Themes are sentences describing the subjects of research of firms or of research organizations. They can be identified in any text, be more disaggregated and provide a subtler representation of knowledge but they are more laborious to identify. The two types of knowledge units are thus not exact substitutes but each of them has both advantages and disadvantages.

At the level of the firm the knowledge base (KB) can be defined as the collective knowledge that firms can use to achieve their productive objectives. The collective character comes from the interactions between individuals, research units and departments of the same firm or research organization. Such interactions are specific to each organization and can be expected to lead to a different knowledge time path even in the case in which the initial competencies of all the persons employed were the same. When we want to study the knowledge base of an industrial sector or of a field of science such collective character of course includes inter organizational interactions.

The KB of a firm can be mapped by identifying the units of knowledge composing it and by their connections or links. As previously explained units can be either technological classes or themes. Connections are determined by the joint utilization of the units in particular texts, be they patents, papers or something else. For example, if we use technological classes the connections are given by the co-occurrence of different classes in the patents used, and the frequency of co-occurrence can be interpreted as a measure of the strength of the link. In this way we can construct visual maps of the KB of a firm and follow the evolution of such KB in the course of time. These maps of the KB can be considered a representation of the brain of the firm.

Even in a knowledge based economy firms are not predominantly knowledge producers but use knowledge in order to achieve new products and services with which they compete. We can find out whether knowledge production is determinant of firm performance by measuring a number of properties of their KB and by using these measures as independent variables in econometric studies in which some measure of firm performance is the dependent variable. Amongst these properties the most important ones are: the *coherence* (COH) of a KB, its *differentiation* (VAR), the *similarity* of two KBs and its converse, the *cognitive distance* (CD) between two KBs. We can expect coherence to fall as firms try to internalise the new type of knowledge constituting for them a discontinuity. However, COH can be expected to start growing again later as the firm increases the fraction of its KB constituted by the new knowledge and learns how to use and integrate the new concepts. In our paper the differentiation of the KB is measured by its variety. We distinguish two types of variety, related and unrelated, and measure them by the informational entropy function. Related variety measures the extent of differentiation at a lower level of aggregation, for example that of a group. Unrelated variety measures the extent of differentiation at a higher level of aggregation at which different groups can be expected to differ considerably amongst themselves. As we will see later, this distinction is very useful in interpreting the evolution of the KBs of our KISs. Cognitive distance is an inverse measure of the similarity of the KBs of our sectors. Thus, we can measure the cognitive distance of the KB of one of out KISs at different times or the CD between the KB of two firms at a given time. In this paper we measure the CD between the KB of each of our sectors at different times. Thus we can map the evolution of the CD for each of our KISs in the period 1981-2005. CD can be considered an approximate measure of the extent of discontinuity in the knowledge used by firms or organisations. We expect cognitive distance to increase very rapidly at the emergence of a discontinuity and subsequently to grow at a lower rate or fall as firms and research organisations improve their understanding of the new type of knowledge. The time paths of these three properties of the KB (COH, VAR, CD) are also likely to be related. Thus, when at a discontinuity a firm begins to use a new type of knowledge, radically different from the ones it previously used, we can expect:

(i) the coherence of the KB to fall in the short run as CD grows since firms are likely to have difficulties in learning and integrating particularly unfamiliar concepts and to start recovering as the new knowledge becomes better connected to the old parts of the KB;
(ii) the differentiation of the KB of the firm to grow as a new type of knowledge starts differentiating after its initial creation;
(iii) the rate of fall of the coherence to be proportional to the rate of growth in the differentiation of the KB[2].

As we will see later, these expectations are still too course and the distinction between related and unrelated variety will provide greater subtlety.

In this paper we map and measure the KB of sectors rather than of firms. As previously pointed out, in this case the KB we map depends on inter-individual and inter-organizational interactions both at the intra- and at the inter-firm level. In this case we can expect to find patterns of evolution reflecting the behaviour of the average or representative firm. Since the sector is a population of broadly comparable firms to have a complete representation of it we would need to measure the distribution of the properties of the KB within the population.

On the basis of the previous considerations we can now formulate the following three propositions:

P1: The emergence of a discontinuity in a type of knowledge suitable to become the future knowledge base of a KIS leads to the sequence of the two periods of random search first occurring in the exploration phase, and of organized search later in the exploitation phase.
P2: During the random search period KB variety rises, KB coherence falls and the cognitive distance between the previous KB of a KIS and the new emerging knowledge rises. During the organized search period the rate of growth of variety falls, KB coherence rises and the cognitive distance between the previous KB of a KIS and the new emerging knowledge falls.
P3: The higher the rate of increase over time in variety and in cognitive distance, and the higher the decrease over time in coherence in the knowledge base, the more persistent the period of random screening, i.e. the less established the organized search period.

If these propositions hold true, then we should expect to observe the following implications in our data:
- In our KISs, there should be a series of initial stages of industry life cycle where exploration of new possibilities takes place, materialized by the experimentation of new IPC classes. Then should follow more mature stages where the exploitation of old certainties predominates, namely with the recombination of existing IPC classes.

[2] Studies of this type have shown that the coherence and the differentiation of the KB of pharmaceutical firms affect both their technological (Nesta, Saviotti, 2005) and their stock market performance (Nesta, Saviotti, 2006). The role played by knowledge creation and utilization in firm performance varies according to sectors but it is not unique to any sector (Nesta, 2008).

- In the initial stages and the mature stages, variety and cognitive distance should evolve in the same direction, while coherence should evolve in the opposite direction.
- The appropriate balance between exploration in the initial phases and exploitation in the mature phase may be difficult to find, and KISs may be confronted to different paths of evolution.

The behaviour and performance of KISs is affected by the emergence of new types of knowledge qualitatively different from those firms in the given KIS were previously using. When such a discontinuity occurs the internalisation by firms in a KIS of a new type of external knowledge which has become a promising source of industrial applications has a number of important organizational implications. For example, such a process would involve the replacement of a very large number of researchers whose competencies lie in the old KB with researchers specialized in the new knowledge. This is a complex process likely to require considerable time and to be particularly difficult in the early phases of the new technology when the relevant competencies are still rare. Furthermore, the development of new type of knowledge is likely to follow a systematic pattern moving away from the early revolutionary period of a new paradigm and towards the more predictable phase of normal science.

3 Measurement of the Knowledge Base

The purpose of our analysis consists of the exploration of the evolution of the properties of the knowledge base, with particular emphasis on the issues of variety, complementarity and similarity. It must be stressed that to introduce some rigidities in the national technological portfolios, and to compensate for intrinsic volatility of patenting behaviour, each patent application is meant to last five years. Let us consider them in further detail:

1) First, we decided to measure technological variety by using the information entropy index. Entropy measures the degree of disorder or randomness of the system, so that systems characterized by high entropy will also be characterized by a high degree of uncertainty (Saviotti, 1988).

   Differently from common measures of variety and concentration, the information entropy has some interesting properties (Frenken and Nuvolari, 2004). An important feature of the entropy measure is its multidimensional extension. Consider a pair of events (X_l, Y_j), and the probability of co-occurrence of both of them p_{lj}. A two dimensional total variety (*TV*) measure can be expressed as follows:

$$TV \equiv H(X,Y) = \sum_l \sum_j p_{lj} \log_2 \left(\frac{1}{p_{lj}} \right) \tag{2}$$

   If one considers p_{lj} to be the probability that two technological classes l and j co-occur within the same patent, then the measure of multidimensional entropy focuses on the variety of co-occurrences of technological classes within regional patents applications.

   Moreover, the total index can be decomposed in a “within” and a “between” part anytime the events to be investigated can be aggregated in a smaller numbers of subsets. Within-entropy measures the average degree of disorder or variety within the subsets, while between-entropy focuses on the subsets measuring the variety across them. Frenken et al. (2007) refer to between- and within- group entropy respectively as unrelated and related variety.

It can be easily shown that the decomposition theorem holds also for the multidimensional case. Hence if one allows $l \in S_g$ and $j \in S_z$ (g = 1,…,G; z = 1,…, Z), we can rewrite *H(X,Y)* as follows:

$$TV = H_Q + \sum_{g=1}^{G}\sum_{z=1}^{Z} P_{gz} H_{gz} \qquad (3)$$

Where the first term of the right-hand-side is the between-entropy and the second term is the (weighted) within-entropy. In particular:

$$UTV \equiv H_Q = \sum_{g=1}^{G}\sum_{z=1}^{Z} P_{gz} \log_2 \frac{1}{P_{gz}} \qquad (4)$$

$$RTV \equiv \sum_{g=1}^{G}\sum_{z=1}^{Z} P_{gz} H_{gz} \qquad (5)$$

$$P_{gz} = \sum_{l \in S_g}\sum_{j \in S_Z} p_{lj}$$

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

We can therefore refer to between- and within-entropy respectively as *unrelated technological variety* (*UTV*) and *related technological variety* (*RTV*), while total information entropy is referred to as *general technological variety*.

2) Secondly, we need a measure of cognitive distance able to express the dissimilarities amongst different types of knowledge. A useful index of distance can be derived from the measure of *technological proximity*. Originally proposed by Jaffe (1986 and 1989), who investigated the proximity of firms' technological portfolios. Subsequently Breschi et al. (2003) adapted the index in order to measure the proximity, or relatedness, between two technologies. The idea is that each firm is characterized by a vector *V* of the *k* technologies that occur in its patents. Knowledge similarity can first be calculated for a pair of technologies *l* and *j* as the angular separation or uncentred correlation of the vectors V_{lk} and V_{jk}. The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{k=1}^{n} V_{lk} V_{jk}}{\sqrt{\sum_{k=1}^{n} V_{lk}^2}\sqrt{\sum_{k=1}^{n} V_{jk}^2}} \qquad (6)$$

The idea underlying the calculation of this index is that two technologies j and l are similar to the extent that they co-occur with a third technology k. The cognitive distance between j and l is the complement of their index of the similarity:

$$d_{lj} = 1 - S_{lj} \qquad (7)$$

Once the index is calculated for all possible pairs, it needs to be aggregated at the industry level to obtain a synthetic index of technological distance. This can be done in two steps. First of all one can compute the weighted average distance of technology l, i.e. the average distance of l from all other technologies.

$$WAD_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jit}}{\sum_{j \neq l} P_{jit}} \tag{8}$$

Where P_j is the number of patents in which the technology j is observed. Now the average cognitive distance at time t is obtained as follows:

$$CD_t = \sum_l WAD_{lit} \times \frac{P_{lit}}{\sum_l P_{lit}} \tag{9}$$

3) Cognitive distance measures the degree of dissimilarity among technologies. We expect it to provide us with an indication of the difficulty, or cost a firm has to face to learn a new type of knowledge. Typically a firm needs to combine, or integrate, many different pieces of knowledge to produce a marketable output. Thus, in order to be competitive a firm not only needs to learn new 'external' knowledge but it needs to learn to combine it with other, new and old, pieces of knowledge. We can say that a knowledge base in which different pieces of knowledge are well combined, or integrated, is a coherent knowledge base. Such technologies are by definition complementary in that they are jointly required to obtain a given outcome For this reason, we turned to calculate the *coherence* (*R*) of the knowledge base, defined as the average relatedness of any technology randomly chosen within the sector with respect to any other technology (Nesta and Saviotti, 2005 and 2006; Nesta, 2008).

To yield the knowledge coherence index, a number of steps are required. In what follows we will describe how to obtain the index at the sector level. First of all, one should calculate the weighted average relatedness WAR_l of technology l with respect to all other technologies present within the sector. Such a measure builds upon the measure of technological relatedness τ_{lj} (see Nesta and Saviotti, 2005, for details). Following Teece et al. (1994), WAR_l is defined as the degree to which technology l is related to all other technologies $j \neq l$ in the sector, weighted by patent count P_{jt}:

$$WAR_{lt} = \frac{\sum_{j \neq l} \tau_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \tag{10}$$

Finally the coherence of knowledge base within the sector is defined as weighted average of the WAR_{lt} measure:

$$R_t = \sum_{l \neq j} WAR_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{11}$$

It is worth stressing that such index implemented by analysing co-occurrences of technological classes within patent applications, measures the degree to which the services rendered by the co-occurring technologies are complementary one another.

The relatedness measure τ_{lj} indicates indeed that the utilization of technology *l* implies that of technology *j* in order to perform specific functions that are not reducible to their independent use. This makes the coherence index appropriate for the purposes of this study.

4 The Data

The information concerning patent applications required to test the working hypotheses formulated in Section 2 has been obtained from the Espacenet data base provided by the European Patent Office[3]. The initial dataset consisted of 2,659,301 items, including both EU and Worldwide applications, over the period 1978 – 2005. The analysis thus focuses on three subsets of patent applications, identified by merging the classifications set up by the OECD and by the French *Observatoire des Sciences et des Techniques*. We adopted these classifications to establish some tentative boundaries for our KISs, although we will realize later in this paper that in some cases these classifications leave some important classes out.

Our search strategy is based on queries reporting the IPC classes that define each KISs under study, namely biotechnology, telecommunications and electronics. Taking into account these elements, it resulted that the biotechnology sector includes 11 IPC classes, the telecommunications sector is made up of 16 IPC classes and the electronics sector consists of 30 IPC classes (see Appendix 1)[4].

Table 1 reports the count of patent applications in each sector and the share in the whole dataset. Although the biotechnology sector is defined by the lowest number of classes, its share in the overall dataset is the highest (12.08%), while the sector gathering the highest number of classes, i.e. electronics, represents the lowest share (1.81%). The telecommunications sector occupies an intermediate position in both the number of classes and its share in the dataset.

INSERT TABLE 1 ABOUT HERE

One can reasonably assume that the dynamics of knowledge production in KISs is marked by important specificities. In this perspective, one immediate (and potentially obvious) specificity is that knowledge production in KISs is likely to be higher than in other sectors.

Although at a first glance our KISs seem to share a common growth pattern, the behaviour of biotechnology is quite different from that of telecommunications and of electronics. The number of patents in biotechnology is about twice as large as in telecommunications and about three times as large as in electronics. Furthermore, the rate of growth of patents in biotechnology seems to be more evenly distributed during the period studied than in telecommunications.

---

[3] We consider thus patent applications as the best indicator of firms knowledge bases, though the usual caveats mentioned in the literature may apply. We use these data to map the frequency of co-occurrences of technological classes within patents and to calculate a number of indexes, i.e. information entropy used to measure related and unrelated variety, knowledge coherence and cognitive distance.

[4] Though the use of IPC classes to define sectors' boundaries may present some drawbacks, as they are function-oriented (Corrocher et al., 2007), the merging of two classifications allows our study to be much more inclusive than many other studies, and reduce the risk of neglecting important classes.

Figure 1 shows the dynamics of technological classes in each sector. Since the dynamics of technological differentiation in the three KISs is influenced by the dynamics of the patent stock, we show the 5-years moving average of classes counts.

INSERT FIGURE 1 ABOUT HERE

The number of classes can be interpreted as an approximate measure of the differentiation, or scope, of each sector's knowledge base. It is interesting to note that within the same sector the differentiation of the knowledge base rises with the stock of patents. However, the influence of the stock on the differentiation of the knowledge base differs for the three sectors. For example, the electronics sector is characterized by a higher number of technological classes than either biotechnology or telecommunications, although its yearly number of patent applications is considerably lower.

5 Empirical results

5.1 Random search versus organized search

The first aspect that we investigated in the results of our calculations was the presence in our three KIS of a transition from random to the organised search. To test the existence of this transition we constructed a co-occurrence matrix of the technologies. used in the patents awarded to the three KIS in our data base. Each patent is classified according to a primary and to a number of secondary classes. Such matrices are constructed by assigning frequencies to the couples of IPC classes occurring together within If the transition from random to organised search occurs we expect a declining fraction of the off diagonal cells to contain a growing share of the overall frequency of co-occurring technologies. In other words, the transition from random to organised search involves a process of *concentration* of the technological choices made in the patents. In a graphic representation of the co-occurrence matrix (Figs A1-A4 Appendix) this phenomenon would be revealed by a growing share of few and higher peaks amongst those representing all the possible technological combinations. We tried to document the existence of the transition from random to organised search by calculating the Gini coefficient for technological co-occurrences[5], starting from the matrix of relative frequency of co-occurrence of technological classes, according to equation (1). The results of these calculations are reported in Figure 2.

INSERT FIGURE 2 ABOUT HERE

There we can see that the Gini coefficient of biotechnology is substantially higher than that of telecommunications and of electronics and hardly changes during the period of observation; the coefficients of telecommunications are higher than those of electronics; the coefficients of both telecommunications and electronics grow during at least a part of the period of observation. The higher relative value of its Gini coefficient indicates that biotechnology had reached a substantially higher level of concentration of technological combinations than

[5] We have calculated the relative Gini concentration index according to the formula $G = \frac{2}{(n-1)}\sum_{i=0}^{n}(P_i - Q_i)$, where *i* refers to the n-th technological co-occurrence, *Q* the observed share of each couple and *P* the share each couple would have had if the distribution would have been equiprobable.

telecommunications and of electronics at the beginning of the period of observation but that this level of concentration hardly changed afterwards. On the other hand, the level of technological concentration of telecommunications and electronics increases starting from 1987 for telecommunications and from 1990 for electronics. To what extent these results confirm or disproof the existence of the transition from random to organised search is not clear without making reference to the measures of variety, coherence and cognitive distance to be shown next. For the time being we can only point to a problem which will affect all our interpretations, namely the duration of our period of observation. In order to be able to test the presence of a transition we would need to cover a period of time starting before the transition and ending after it. We know that the first industrial applications of biotechnology started in the early to mid 1970s but our observations begin in 1981. Thus, we cannot decide whether in 1981 the transition had already occurred for biotechnology or whether biotechnology underwent no such transition and it always had such a high level of technological concentration. Based on their Gini coefficients telecommunications and electronics could have undergone the transition from random to organised search. In the end, whether or not we find evidence in favour of the transition, it is quite clear that the three KIS behave differently. Whatever the interpretation we attach to these findings we can consider the Gini coefficient an index of technological concentration and we can treat it as a relevant property of the knowledge base of technologies or industrial sectors.

5.2 The Evolution of KBs in KISs

The measures of variety, coherence and cognitive distance are here applied to investigate the patterns of evolution of knowledge bases in three broad sectors: biotechnology, telecommunication and electronics.

5.2.1 Biotechnology

Figure 3 shows the evolution of variety (a), coherence (b) and cognitive distance (c) for the biotechnology sector. Moreover, the rate of growth of variety falls for most of the period of observation until it becomes constant from the early 1990s, with the possible exception of the mid 1980s. In 1985 the rate of growth of variety starts rising in correspondence with the overtaking of unrelated variety by related variety. The distinction between unrelated and related variety was introduced by Frenken et al. (2007) to measure the output variety of different regions of the Netherlands. In our case while in the early 1980s the unrelated variety was higher than the related, the situation was reversed starting from 1985. This would suggest that, while in the very early phases of the emergence of modern biotechnology most of the new knowledge was coming from outside the knowledge base previously used, starting from 1985 internal (to the sector) sources of knowledge differentiation became more prominent. However, it must be observed that starting from the mid 1990s a trend began to the convergence of related and unrelated variety. This trend is likely to be caused by the emergence of a second generation of biotechnology linked to bioinformatics, a new type of competence coming from a discipline different from biology. Further evidence about the relationship among the variables can be found in Table 2a, where the Spearman's correlation coefficient enables to appreciate both the sign and the strength of relations.

INSERT TABLE 2 ABOUT HERE

The distinction between related and unrelated variety is based on the assumption that any pair of entities included in the former are in general more closely related, or more similar, than any pair of entities included in the latter. This assumption is reasonable when a given type of entity (patents, industrial sectors, trade categories etc.) is organised according to a hierarchical classification. In this case each class at a given level of aggregation contains 'smaller' classes, which in turn contain even 'smaller' classes. The term small here corresponds to a low level of aggregation. We can then reasonably expect that the average pair of entities at a given level of aggregation will be more similar than the average pair of entities at a higher level of aggregation. Thus, what we call related variety is measured at a lower level of aggregation than unrelated variety. This distinction is important because in view of the previous discussion we can expect the dominance of unrelated (or inter-group) variety to lead to a greater extent of discontinuity in knowledge than the dominance of related (or intra-group) variety.

INSERT FIGURE 3 ABOUT HERE

Figure 3 b) shows the dynamics of knowledge coherence for the biotechnology sector. In this case as well as in all the other measures of properties of the knowledge base we can distinguish within the overall changes a trend and superimposed deviations. The deviations are probably due to combination of real events affecting the dynamics of knowledge and of noise due to the quality of the data. Thus, we cannot expect all the deviations to be easily interpretable. Both variety and coherence show an overall positive trend accompanied by superimposed deviations. In particular, there are two periods of fast rise in knowledge coherence, beginning in 1982 and in 1995 respectively. The first of these deviations from the trend seems to be closely related to the ratio of related to unrelated variety. When unrelated variety is greater than the related one, in the period 1981-1982, the coherence index falls. It then begins to increase in 1983 when related variety overtakes unrelated variety. The subsequent rise in 1997 cannot be explained in the same way. However, it can be observed that the two rises in knowledge coherence seem to coincide with the onset of the absorption of two different generations of biotechnology, based on recombinant DNA and on genomics respectively, by incumbent firms (Saviotti, Catherine, 2008). The transition between the two generations led to a discontinuity in the pattern of inter-firm alliances: within each generation the number of alliances followed a lifecycle, increasing first, reaching a maximum and then declining. The competencies required in the two generations differed as bioinformatics acquired a in the sequencing of genomes.

Taking this into account we can interpret the overall rising trend in knowledge coherence as due to the growing relative similarity, or low cognitive distance, of the new types of knowledge which incumbent firms needed to learn. The deviations with respect to the trend could be explained by the emergence of new generations of biotechnology and/or by the ratio of intra to inter group variety. As a new generation of biotechnology emerges the overall trend is not reversed but deviations can occur due to the however limited cognitive distance that the new generation introduces. This line of explanation is not incompatible with the one based on the ratio of related to unrelated variety. We can assume changes in related variety to involve a more limited change in coherence than those in unrelated variety because the former can be obtained by recombination and differentiation of the same concepts while the latter are more likely to involve the introduction of completely new concepts. In other words, a rise in related variety is likely to involve a lower extent of knowledge discontinuity than an equivalent rise in unrelated variety and to lead to lower fall in coherence. Conversely we can expect changes of generation within one technology (e.g. biotechnology) to raise the ratio related/unrelated while the emergence of a completely new technology can be expected to lower the same ratio. However, in some cases the situation can be more complex. In this context the transition between the two generations of biotechnology involved two contrasting trends: the second generation shared the same basic biological concepts with the first generation but required the use of competencies and concepts in bioinformatics which were new to biologists and which came from another discipline. We can expect the first trend to raise both related variety and coherence and the second to reduce both of them. What we observe is then the result of a trade-off between the two trends described above. This interpretation is compatible with (i) the tendency to the convergence of related and unrelated variety beginning in the mid 1990s and (ii) the slow down in the rate of growth of coherence between 1988 and 1996 followed by a rise in coherence beginning in 1997, which could be due to the maturation of the second generation of biotechnology.

The evolution of cognitive distance is reported in the bottom part (c) of Fig 3. Even in this case we can distinguish an overall trend from the deviations with respect to it. The evidence for the biotechnology sector is very consistent with the measures of variety and knowledge coherence. The distance index indeed decreases dramatically in the early years of the period we observed. Although with some cyclical fluctuations, it keeps falling until the first half of the 1990s. Then it remains almost constant with the possibility of a very limited rise.

In summary, in the biotechnology sector there has been a growing knowledge differentiation, represented by the growth in variety, accompanied by a trend towards increasing knowledge coherence and towards falling cognitive distance. The rate of growth of variety and of coherence as well as the rate of fall of cognitive distance decrease in the course of time. These broad trends have been combined with a changing ratio of related to unrelated variety and with fluctuations with respect to the trend of both knowledge coherence and of cognitive distance. If we take into account the deviations from the trend in coherence we can see that coherence was falling at the beginning of the period of observation and that it started to grow when related variety overcame unrelated variety. We can interpret these events as follows:

- A drastic fall in the coherence of the knowledge base of biotechnology using firms, which we expect to have started in the early 1970s before the beginning of our period of observation, and continuing until 1983. This was due to the incorporation of completely new elements of knowledge in the KB of biotechnology using firms and organisations
- A subsequent recovery of coherence due to the growing ratio of related to unrelated variety and to the learning effects that occurred in the biotechnology using firms and organizations allowing them to improve their ability to integrate the new knowledge in their KB.
- Subsequent variations of coherence superimposed on a growing trend were due to the emergence of a second generation of biotechnology involving the addition of new types of new types of knowledge (bioinformatics) to the basic biological concepts introduced during the first generation.

It seems clear that the inverse correlation between variety growth and cognitive distance which we could have expected ex-ante does not always occur. In particular, it is clear that related and unrelated variety do not have the same impact on cognitive distance. Since related variety can grow by differentiation in the vicinity of the previous knowledge of firms or organisations, it does not necessarily lead to a rise in cognitive distance. In fact, a rise in related variety combined with the learning effects of firms which allow them to integrate different pieces of knowledge can even be compatible with a fall in cognitive distance. The distinction between related and unrelated variety turns out to be as fruitful in the study of structural change in knowledge as it is in the study of structural change in economic systems (see Frenken et al, 2007; Saviotti, Frenken 2008).

Our findings so far are thus not incompatible with propositions P1-P3. Although the expected fall in knowledge coherence when the new type of knowledge emerged did not occur, the deviations with respect to the trend bear a close relationship to both the emergence of new generations of biotechnology and to the changing ratio of related to unrelated variety. Propositions P1-P3 were initially formulated without taking into account the distinction between related and unrelated variety and will need to be modified accordingly.

5.2.2 Telecommunications

The evidence about Telecommunications is considerably different from biotechnology (Figure 4 a, b, c). Except for the very early years (1981-1982) during which all types of variety rise very rapidly, unrelated variety remains virtually constant and the growth in total variety is almost exclusively determined by related variety. Interestingly between 1991 and 1995 related variety seems to be undergoing a transition which increases substantially its rate of growth. Following the previous reasoning this behaviour could be explained if radically new concepts had been introduced into telecommunications before the beginning of our period of observation and if all the following rise in variety had taken place by 'local' differentiation, obtained for example by recombination of already known concepts or eventually by new forms of exploitation, for example new types of industrial applications. This can be explained looking back into the history of the sector (Fransman, 2002, 2004, 2006, 2007). Since the early 1980s – and even before – the national telecoms operators (at that time monopolists) were leaders in researching and designing the equipment and other technologies, thanks to their research laboratories. These laboratories (France Telecom’s CNET, Telecom Italia’s CSELT, BT’s Martlesham Laboratories in Europe, or AT&T’s Bell Laboratories or NTT’s Electrical Communications Laboratories overseas) regrouped researchers that won lots of Nobel prizes and were at the origins of the Internet and mobile technological development. At a time of liberalization, these laboratories played a key role when newcomers entered. Because of their existence, most of the new specialist providers (Nokia, Ericsson, Cisco, etc.) could progressively supply the latest technologies to all new companies who could pay for it, since they found commercial opportunities to the knowledge incorporated in the patents that research laboratories have registered 10 or 15 years before. In this perspective, the specialist suppliers can be considered as the innovators, while the research laboratories are inventors and repositories of knowledge.

Coherence falls slowly between 1981 and 1991 and then begins to rise when the rate of growth of related variety starts to increase. Even in this case we see that a rise in related variety is not incompatible with a rise in coherence. Slightly more difficult to explain is the slow fall in coherence in the period 1981-1991. This can basically be related to the shift in technological paradigms, from the old one related to the circuit-switched technologies providing basic services like telephony and fax to the new one termed as packet-switched technologies and providing advanced services like the Internet, Video-conferencing, Video on Demand, Voice on IP, etc. (Fransman and Krafft, 2002; Krafft, 2004, 2007; Krafft and Salies, 2008). The important point to notice here is that a rise in related variety does not necessarily lead to a fall in coherence but is compatible with both a rise or a limited fall.

For telecommunications cognitive distance is almost constant or at most falls very gently, with very pronounced deviations from the trend. Even in this case a rise in related variety does not necessarily involve a rise in cognitive distance. In Table 2b the Spearman's correlation coefficient among variables provides an overall synthesis of the relationships linking coherence, variety and cognitive distance.

Together these results can be interpreted by saying that the impact of a knowledge discontinuity giving rise to modern telecommunications, namely the convergence with IT and the transition from analogic-electromechanical to digital-electronic technology, is likely to have started long before the beginning of our period of observation. One could then expect rises in cognitive distance and falls in coherence to have occurred during this early period. We could then expect of the developments occurring during our period of observation to have been, incremental improvements of the knowledge base aimed at creating new industrial applications based on the based on concepts which had already become part of the knowledge base of telecommunications firms.

INSERT FIGURE 4 ABOUT HERE

Compared to biotechnology, telecommunications shows a less smooth trend towards growing knowledge variety and a larger departure of intra- from inter-group knowledge variety. This indicates that during the period studied new forms of knowledge being used in telecommunications were increasingly similar to those already present within the sector. This interpretation is confirmed by the almost constant value of cognitive distance. Furthermore, the relative rise in intra-group knowledge variety seems to indicate a progressive focusing of new forms of knowledge inside the technology. These trends could be interpreted as a growing weight of exploitation relative to exploration in the research activities in telecommunications.

5.2.3 Electronics

With respect to the two previous cases electronics shows another and altogether considerably different development pattern. Unrelated and related variety keep growing all the time while unrelated variety is constantly higher than related variety. Coherence falls all the time while cognitive distance shows a decreasing trend, but at a definitely slower rate than in the other two sectors. This development pattern can be interpreted as characterized by a persistent emergence of considerably new concepts during the whole period of observation. Thus, in this case the duration of the period in which the emergence of completely new concepts occurred seem to have lasted considerably longer than for both biotechnology and telecommunications. The likely origin of this difference is the much wider range of applications of electronics with respect to both biotechnology and telecommunications. This trend could be interpreted as a continued parallel development of exploration and exploitation related research activities. We could also say that electronics is a more general purpose technology than either telecommunications or biotechnology. Moreover, it must be noted that the origins of the electronics sector are much more remote than for biotechnology and telecommunications, as they can be dated back to the invention of the electric light. A closer look at the IPC classes that are included in this sector (see the Appendix) indeed reveals that a considerable number point to light- or energy-related technologies. Only a few technological classes are related to the most recent evolution of the electronics sectors, like H05K (printed circuits) and H04M (telephonic communication). Post-war technological developments have indeed been induced by the need to create and improve devices for long distance telephone communications. Thus, the electronic revolution started with the invention of the transistor in the late 1940s. The evolution of this technology then led to the invention of the integrated circuit in the late 1950s and eventually to that of the microprocessor by Intel in 1971. This last technology has brought profound changes in both the electronics sector itself, and in many other user sectors, as it was a powerful and general purpose solution to many diverse applications. The conditions for the creation and the massive diffusion of minicomputers were therefore set. Computing technology began to be applied to an unprecedented number and diversity of uses, such as telecommunications, banking, car production, etc. This generated an increasing recombination of electronics knowledge with technologies developed in other sectors, even if loosely related, which could have been dramatically improved by the application of computing power (Mowery and Rosenberg, 1998; Mowery and Nelson, 1999; Bresnahan and Malerba, 1999). The increasing convergence with telecommunications, which led to the creation of information and communication technologies (ICTs), is reflected in the evidence of the first half of the 1990s, when knowledge coherence appears to rise while cognitive distance is stable. However, the second half of the 1990s provides evidence of a slightly fall in coherence and of an increase in cognitive distance, which may be ascribed to the emergence of an exploitation pattern related to the Internet revolution.

In electronics general variety keeps increasing until 1998, when it basically stabilizes (Figure 8). However, interestingly, in this case total knowledge variety is led by unrelated variety, which always has a higher weight than the related one. This can be interpreted if we consider that electronics is at the interface of different applications developed in several industries. Gradually over time, very different pieces of knowledge have to be reassembled, due to the demand of client sectors (including telecommunications, automobile, banks, computers, medical instrumentation, etc.). Thus, electronics really appears as a general purpose technology the pervasiveness of which is more and more prominent over time.

INSERT FIGURE 5 ABOUT HERE

Not surprisingly, the trend of knowledge coherence falls all the times. The fluctuations show falls beginning in 1981 and in 1989 and rises beginning in 1985 and in 1993, presumably because of the increasing number of applications of electronics. In this case the overall trend with respect to biotechnology and to telecommunications can be explained by the higher relative value of inter-group knowledge variety. This implies a higher cognitive distance, as confirmed by the bottom diagram. On the whole, the evidence about the electronics sector appears to be fairly different from that of biotechnologies and telecommunications. Such differences may be further appreciated by looking at the last group of correlation coefficients provided in Table 2c.

5.2.4. Summing up

The Table below (Table 3) is intended to sum up the results we obtained so far. It regroups the propositions we made in Section 2, their major expected implications, and what observed in our three KISs

INSERT TABLE 3 ABOUT HERE

6 Discussion and conclusions

In this paper we studied the dynamics of knowledge generation in three knowledge intensive sectors (KIS), biotechnology, telecommunications and electronics. We mapped the knowledge base of these three KISs by means of the patents awarded in them by the European Patent Office (EPO) during the period 1981-2002. We did not distinguish the different types of economic actors to which the patents were given but considered each sector as a whole. For each sector we measured four properties of the knowledge base: technological concentration, variety, coherence and cognitive distance. Within variety, a variable which measures the extent of differentiation of the knowledge base of each sector, we distinguished related from unrelated variety. The former is the variety which we can measure within a group of entities at a lower level of aggregation, the latter variety which can be measured at a higher level of aggregation. The two measures have a different meaning in that we can expect the average pair of entities constituting a group at the lower level of aggregation to have a higher degree of similarity than the average pair of entities in two different groups, or at a higher level of aggregation. This distinction is interesting because we can expect an increase in unrelated variety to correspond to a more radical change in knowledge, and thus to a greater extent of knowledge discontinuity, than an equivalent increase in related variety. From our results, and considering the ratio related/unrelated variety, it appears that biotechnology and telecommunications progressively enter into a more mature phase of development. It appears that biotechnology and telecommunications have already entered, although at different times,a more mature phase in which exploitation related activities tend to to grow with respect to exploration related ones. As we previously pointed out, the concepts of exploration and exploitation are very useful although not analytically accurate. The properties of the knowledge base that we measure in our paper provide a means to make these concepts more analytical. Thus we can expect certain regular relationships to exist amongst the properties we measure and between these properties and exploration and exploitation. For example, we can expect exploration to be associated with growing technological concentration, with growing overall variety, with a high or growing ratio RTV/UTV, with low or falling coherence, with high or growing cognitive distance. A complete representation of these expected relationships is given in Table 4.

INSERT TABLE 4 ABOUT HERE

As we can see in table 4, the correspondence between the properties of the KB allows multiple patterns to occur. For example, a high ratio RTV/UTV which we expect to find in the exploitation phase is compatible with both a high, constant or growing coherence. Thus, the relationships shown in Table 4 only exclude particular patterns of correspondences but allow more than one pattern to correspond to each of the two phases. This means that the exploration/exploitation dichotomy is a simple and useful one but that our representation of the knowledge base in terms of more objective and measurable properties is a more accurate one. In spite of the lack of complete correspondence between the phases of exploration and exploitation and the expected values of our properties the availability of such properties clarifies the meaning of exploration and exploitation and greatly helps in giving the two phases an operational use.

The technological concentration of a knowledge base measures the extent to which a small number of technological combinations accounts for a large share of the technological choices made in the research activities constructing the knowledge base of the KIS we are studying. It must be observed that the meaning of technological concentration used here is very similar to that of industrial concentration or of the distribution of income. In all these cases an index of concentration measures the extent to which the distribution of a property within a population becomes skewed or asymmetrical. Furthermore, technological concentration here seems to increase spontaneously in the course of time similarly to the spontaneous evolution towards oligopolies occurring in many industrial sectors. The coherence of a knowledge base measures the extent to which a given firm or research organization can combine, or integrate, the different pieces of knowledge which are required to produce a given product or service. The cognitive distance measures the dissimilarity of two knowledge bases. It can be used to compare the knowledge bases of two different firms at a given time or to calculate the changes in the knowledge base of the same firm in the course of time. These three properties can be expected to undergo systematic variations when a firm or research organization needs to learn a new and unfamiliar type of knowledge. The above properties of the knowledge base are not independent. At the beginning of our study we expected an increase in variety due to the emergence of a completely new type of knowledge to lead to a fall in coherence and to a rise in cognitive distance. In a previous and less detailed study (Grebel, Krafft, Saviotti, 2006) we had found evidence of a common trend in knowledge generation which we described as the transition from 'random search' to 'organised search'. In the early phases of the emergence of a knowledge discontinuity, during the random search period, firms and research organisations start exploring a new region of knowledge space without knowing precisely what trajectories to follow.

The transition to the organised search period occurs as some particularly fruitful research trajectories emerge, which are then followed by the majority of participants. In the present paper we found evidence for this transition in the three sectors, but, by carrying out a more detailed study, we also found evidence of differences amongst the sectors. Thus, for the three sectors variety increased most of the time during our period of observation. However, for both biotechnology and telecommunications, and even more so for the latter, growth in variety was dominated most of the time by related variety while for electronics variety growth was always dominated by unrelated variety. This means that the growth of knowledge in electronics during the period of observation was due to more radical changes in knowledge while in biotechnology and telecommunications the growth of knowledge was mostly due to incremental changes. It is to be observed that in the very early years of our period of observation in biotechnology unrelated variety was greater than related variety and that the dominance of related variety only began in 1983. Unfortunately our patent time series do not cover the whole period which would be required to observe completely the emergence and maturation of a new type of knowledge.

Thus, in biotechnology the research leading to the creation of a new discipline (molecular biology) began in the 1930s and the critical events which catalysed the first industrial applications only occurred in the early to mid 1970s (1972 recombinant DNA, 1975 monoclonal antibodies). In order to adequately study the evolution of knowledge in biotechnology our data would have needed to cover most of the 1970s. Given the limitations of our data for the time being we have to infer what is likely to have happened before the beginning of our period of observation. In the case of biotechnology, based on the very low initial value of both variety and coherence and on the fact that coherence was still falling at the beginning of the period of observation, we expect unrelated variety to have been greater than related variety during all of the 1970s and until 1983. Thus, the 1970s would have been the period when the discontinuity in biotechnological knowledge constituted by the adoption of molecular biology would have first manifested itself and the 1980s the period during which the new knowledge started to be adequately integrated into the knowledge base of biotechnology using firms. We can interpret this transition as being related to the one from exploration to exploitation.

In the case of telecommunications the emergence of a discontinuity is likely to have occurred even earlier than in biotechnology and thus the dominance of related variety is likely to have started before the beginning of our period of observation. Also, it is to be noticed that telecommunications is a sector highly oriented towards applications and that its knowledge base overlaps that of electronics. In fact, the most important recent development in telecommunications is the convergence with information technology (IT) giving rise to ICT and to the so-called new info-communications industry. The critical events underlying the emergence of IT first and of ICT later, the invention of the transistor etc, occurred in the 1950s. Thus, not only these critical events occurred earlier than in biotechnology, but telecommunications received most of its knowledge from another sector and precisely from electronics. Yet, the new info-communications industry is grounded on former knowledge (IP and mobile technologies) which did not find for a long time innovative applications. Only in the 1990s the emergence of packet-switched technologies on which the Internet is based, generated a new set of commercial applications,. While the research laboratories tended to open up new avenues in terms of inventions, the knowledge base of the telecommunications operators was essentially related to the applications of circuit-switched technology, and required a drastic change in competencies to adapt to the new industrial challenges. New technology based firms (IP based, like the new equipment providers) emerged with liberalisation and contributed to the gradual change in the knowledge base, evolving towards a greater coherence of former technological knowledge with more recent market premises. Summarising, knowledge production in telecommunications during our period of observation was largely due to knowledge imported from electronics and IT and highly application oriented. Furthermore, the critical events leading to industrial applications are likely to have occurred before the beginning of our period of observation.

Electronics was considerably different with respect to both biotechnology and telecommunications. In this case unrelated variety always dominated related variety, coherence fell and cognitive distance increased all the time. These patterns can be interpreted as due to the more radical character of the changes in knowledge which keep occurring in electronics as compared to biotechnology and telecommunications. In general we can expect cognitive distance to be a measure of the extent of knowledge discontinuity occurring within a given field. The higher the cognitive distance between the knowledge base of any organisation between two time periods the more difficult it is for the organisation to integrate different pieces of the old and of the new knowledge it needs to use. Thus, a growing cognitive distance is likely to be accompanied by a falling coherence.

A better appreciation of the differences and similarities occurring among the three sectors with respect to the knowledge related variables can be reached by looking at Table 5. Here one can find pairwise mean comparison tests for each variable.

INSERT TABLE 5 ABOUT HERE

The important role of the distinction between related and unrelated variety can be understood here. Since related variety is linked to more incremental types of change and unrelated variety to more radical ones, we can expect coherence to fall and cognitive distance to rise all the times when unrelated variety increases while this does not necessarily occur for a rise in related variety. In this case a process of growing knowledge differentiation can occur based on the set of concepts which were introduced at the very emergence of the discontinuity.

Thus, we have seen that a growth in related variety is compatible with a rise in coherence and with fall in cognitive distance.

Based on the these results we can hypothesize that any new type of knowledge will follow a life cycle beginning with the emergence of a knowledge discontinuity and dependent on the initial value of cognitive distance. The early phases of the life cycle would correspond to (i) low technological concentration (ii) a growth in total variety, (iii) the dominance of unrelated over related variety, (iv) falling coherence, (v) growing cognitive distance. The maturation of the new type of knowledge would entail (i) a rise in technological concentration (ii) a continuation of the growth of total variety, (iii) a shifting dominance of related over unrelated variety, (iv) a slowly falling and later growing coherence, (v) a falling cognitive distance.

The emergence and the maturation of a KIS would here correspond closely to the exploration and exploitation phases. In fact, the properties of the KB which we use in this paper allow us to provide a more analytically accurate representation of the concepts of exploration and exploitation. We have previously seen that there is not a one to one correspondence between values of our KB properties and the phases of exploration and exploitation. Multiple patterns and combinations of these properties can occur within each of the two phases. The use of the four properties of the KB makes the analysis of the KIS richer and more accurate and improves the operational use of the concepts of exploration and exploitation. Furthermore, we have seen that the transition from random to organised search does not occur in a standardised way for all KIS when a knowledge discontinuity emerges. On the contrary, by using our four properties we can measure the extent of such knowledge discontinuity, follow its evolution and see how it affects and is affected by the other properties. Unsurprisingly, we then find that the evolution of each KIS, while being broadly compatible with the transition from random to organised search, presents some significant sectoral specificities. For example, we have seen that the timing of the transition from the initial to the mature phases, the ratio of unrelated to related variety, the overall extent of cognitive distance can vary considerably amongst the sectors studied. Thus, if a knowledge life cycle effectively exists its description must include the factors which can determine the existence, duration and the internal dynamics of the life cycle.

With this paper we have attempted a first exploration of the dynamics of knowledge in KIS. Research of this type is very important as we move towards a knowledge based economy and society since it can create the tools required to represent and measure knowledge. We find the results fascinating but, as it befits the initial exploration of a new field, our results are hardly complete or definitive. The central aspect of our research is the mechanisms whereby new knowledge can be learned by knowledge producing and using organizations. In this paper we provided map of the evolution of knowledge in three KIS. Our findings suggest some general conclusions which will need to be tested and better articulated.

Amongst the problems which arise from this paper and which require additional work there are: (i) the further exploration of the fine structure of knowledge dynamics in each of the KIS studied here, for instance by relying more extensively to monographs or business history analyses, (ii) the comparison with other sectors of different knowledge intensity, for example in view of quantifying more generally the relationships between our three properties, (iii) the impact of these processes of knowledge generation on industrial organization, especially by including in our analysis aspects of entry and exit but also by relating problems of knowledge creation with geographical concerns. We intend to explore all these topics in our future research agenda.

References

Agarwal, R., Gort, M., 2001, "First-Mover Advantage and the Speed of Competitive Entry, 1887-1986", *Journal of Law and Economics*, 44(1), 161-77.

Antonelli, C., 2008, *Localized Technological Change: Towards the Economics of Complexity*, Routledge: London.

Archontopoulos, E., Guellec, D., Stevnsborg, N., Van Pottelsberghe de la Potterie, B., Van Zeebroeck, N., 2007, "When Small Is Beautiful: Measuring the Evolution and Consequences of the Voluminosity of Patent Applications at the EPO", *Information Economics and Policy*, 19(2), 103-32.

Attaran, M., 1985, "Industrial dinversity and economic performance in US areas", *The Annals of Regional Science*, 20, 44-54.

Breschi, S., Lissoni, F., and Malerba, F., 2003, "Knowledge relatedness in firm technological diversification", *Research Policy*, 32, 69-97.

Bresnahan T. F. and Malerba, F., 1999, Industrial dynamics and the evolution of firms' and nations' competitive capabilities in world computer industry, in Mowery, D. C. and Nelson R. R., *Sources of Industrial Leadership. Studies of Seven Industries*, Cambridge, Cambridge University Press, pp.79-133.

Corrocher, N., Malerba, F., Montobbio, F., 2007, "Schumpeterian Patterns of Innovative Activity in the ICT Field", *Research Policy*, 36(3), 418-32.

Dasgupta P., David P., 1994, "Towards a new economics of science". *Research Policy*, 23, 487-521.

Fransman, M. (Ed.), 2006, *Global Broadband Battles: Why the US and Europe Lag While Asia Leads*, Stanford: Stanford University Press.

Fransman, M., 2002, *Telecoms in the Internet Age: From Boom to Bust to…*, Oxford

Fransman, M., 2004, "The telecoms boom and bust 1996-2003 and the role of financial

Fransman, M., 2007, The New ICT Ecosystem: Implications for Europe, Edinburgh: Kokoro.

Fransman, M., and Krafft, J., 2002, "Telecommunications", in W. Lazonick (Ed.) *IEBM Handbook of Economics*, Thomson Learning.

Freeman, C., Soete, L., 1997, The Economics of Industrial Innovation, Third edition. Cambridge: MIT Press.

Frenken, K., Nuvolari, A., 2004, "Entropy Statistics as a Framework to Analyse Technological Evolution", in John Foster and Werner Hölzl (eds), *Applied evolutionary economics and complex systems*, Cheltenham, U.K. and Northampton, Mass.: Elgar.

Frenken, K., von Oort, F., Verburg, T., 2007, "Related Variety, Unrelated Variety and Regional Economic Growth", *Regional Studies*, 41(5), 685-97.

Grebel, T., Krafft, J., Saviotti, P.P., 2006, "On the Life Cycle of Knowledge Intensive Sectors", *Revue de L'OFCE*, june, 63-85

Griliches, Z., 1979, Issues in assessing the contribution of research and development to productivity growth, *The Bell Journal of Economics*, 10, 92-116.

Griliches, Z., 1990, "Patent statistics as economic indicators: a survey", *Journal of Economic Literature*, 28, 1661-1707.

Jaffe, A., 1986, "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value", *American Economic Review*, 76(5), 984-1001.

Jaffe, A., 1989, "Real Effects of Academic Research", *American Economic Review*, 79(5), 957-70.

Klepper, S., 1997, "Industry Life Cycles," *Industrial and Corporate Change*, 6(1), 145-81.

Krafft, J., 2004, "Entry, Exit and Knowledge: Evidence from a Cluster in the Info-communications Industry, *Research Policy*, 33(10), 1687-1706

Krafft, J., 2006, “Industrial Dynamics – Introduction to the Special Issue”, *Revue de l’OFCE*, june, 15-20.
Krafft, J., 2007, “Telecommunications, the Internet and Mr Schumpeter”, in H. Hanusch and A. Pyka (eds) Elgar Companion to Neo-Schumpeterian Economics, Cheltenham: Edward Elgar.
Krafft, J., and Salies, E., 2008, “The cost of switching Internet providers in the broadband industry, or why ADSL has diffused faster than other innovative technologies: Evidence from the French case”, *Research Policy*, 37(4), 706-719.
March, J., 1991, “Exploration and exploitation in organizational learning”, *Organization Science*, 2(1), 71-87.
March, J.G., 1991, Exploration and exploitation in organizational learning, *Organization Science*, 2, 71-87.
Mowery, D. C. and Nelson R. R., 1999, *Sources of Industrial Leadership. Studies of Seven Industries*, Cambridge, Cambridge University Press.
Mowery, D.C. and Rosenberg, N., 1998, *Paths of Innovation: Technological Change in 20th Century America*, Cambridge: Cambridge University Press.
Narin, F., 1994, “Patent bibliometrics”, *Scientometrics*, 30**(**1), 147–155.
Nesta, L., 2008, “Knowledge and productivity in the world’s largest manufacturing corporations”, *Journal of Economic Behavior and Organization*, 67, 886-902.
Nesta, L., and Saviotti, P.P., 2005, “Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry”, *Journal of Industrial Economics*, 53(1), 123-42
Nesta, L., and Saviotti, P.P., 2006, “Firm Knowledge and Market Value in Biotechnology”, *Industrial and Corporate Change*, 15(4), 625-52
OECD, 2005, A framework for biotechnology statistics, http://www.oecd.org
Piscitello, L., 2004, “Corporate Diversification, Coherence and Economic Performance”, *Industrial and Corporate Change*, 13(5), 757-87
Saviotti P.P., Catherine D., Innovation networks in biotechnology, forthcoming in *Handbook of Bioentrepreneurship*, Holger Patzelt, Thomas Brenner, David B. Audretsch (Eds) Springer (2008)
Saviotti, P. P., 1988, "Information, variety and entropy in technoeconomic development," *Research Policy*, 17(2), 89-103.
Saviotti, P., and Frenken, K., 2008, "Export variety and the economic performance of countries", Journal of Evolutionary Economics, 18(2), 201-218.
Saviotti, P.P., 1996, Technological Evolution, Variety and the Economy, Cheltenham: Edward Elgar.
Saviotti, P.P., 2004, Considerations about the production and utilization of knowledge, *Journal of Institutional and Theoretical Economics*, 160, 100-121.
Saviotti, P.P., 2007, On the dynamics of generation and utilisation of knowledge: The local character of knowledge, *Structural Change and Economic Dynamics*, 18, 387-408.
Teece, D. et al., 1994, “Understanding Corporate Coherence: Theory and Evidence”, *Journal of Economic Behavior and Organization*, 23(1), 1-30
Theil, H., 1967, *Economics and Information Theory*, Amsterdam: North Holland, University Press: Oxford.
Verspagen, B., 1997, “Measuring inter-sectoral technology spillovers: estimates for the European and US patent office databases, *Economic System Research*, 9(1), 49-67.

Table 1 – Overall distribution of patent applications across the three sectors

	#	%
Biotechnology	321449	12.08
Telecommunications	115735	4.35
Electronics	47955	1.81

Table 2 – Spearman rank correlation coefficients across variables, by Sector

BIOTECHNOLOGY

	Coherence	Gini index	Cognitive distance	RTV/UTV
Coherence	1			
Gini index	0.4455**	1		
Cognitive distance	-0.9390***	-0.4740**	1	
RTV/UTV	0.4675**	-0.1078	-0.5325***	1

a)

TELECOMMUNICATIONS

	Coherence	Gini index	Cognitive distance	RTV/UTV
Coherence	1			
Gini index	0.4403**	1		
Cognitive distance	-0.303*	-0.8364***	1	
RTV/UTV	0.4221**	0.8961***	-0.8338***	1

b)

ELECTRONICS

	Coherence	Gini index	Cognitive distance	RTV/UTV
Coherence	1			
Gini index	-0.2740	1		
Cognitive distance	0.5974***	-0.5727***	1	
RTV/UTV	-0.64629***	0.5117***	-0.9338***	1

c)

Note: ***: p<0.01; **: p<0.05; *: p<0.1.

Table 3 - Synthesis of empirical results

Propositions	Implications	Empirical Results
P1: Emergence of a discontinuity in knowledge: period of random search in the exploration phase versus period of organized search in the exploitation phase	KISs should evolve from initial stages to more mature stages of industry life cycles	Proposition holds for Biotechnology and Telecommunications. Electronics however persistently in a phase of random search
P2: The random search period is characterized by raising variety and cognitive distance and by falling coherence. The organized search period is characterized by falling variety and cognitive distance and by raising coherence	In the initial stages of development of an industry life cycle, variety and cognitive distance rise, while coherence falls. In the more mature phases of development of an industry life cycle, variety and cognitive distance decline, while coherence increases	Proposition appears not to be robust, since the distinction between related and unrelated variety is important and introduces some complexities. Biotechnology and Telecommunications which over time are characterized by an increasing ratio related/unrelated variety are able to maintain increasing coherence and declining cognitive distance. On the contrary, Electronics where unrelated variety is persistently high is more conform to the proposition and the possible implication is that it has not reached yet a mature phase of development in the industry life cycle
P3: The higher the rate of of increase over time in variety and cognitive distance, the higher the decrease over time in coherence, the more persistent the period of random screening and the less established the organized screening period	The appropriate balance between exploration in the initial phases and exploitation in the mature phases is likely to be difficult to find, and KISs may be confronted to different paths of evolution	Again, proposition appears not to be robust, since the distinction between related and unrelated variety is important and introduces some complexities. When the ratio related/unrelated variety remains low over time (Electronics), proposition holds. When the ratio related/unrelated is higher (Biotechnology and Telecommunications), increasing coherence may be achieved, as well as decreasing or stagnating cognitive

Table 4 - Expected relationships of the properties of the knowledge base during the exploration and exploitation phases

	TC	VAR	RV/UV	COH	CD
Exploration	Low	High	Low	Low	High
Exploitation	High	Falling or	High	High, constant or growing	Low or falling

Note: TC = technological concentration; VAR = overall variety; RV = related variety; UV = unrelated variety; COH = coherence; CD = cognitive distance

Table 5 - Pairwise T-test for equality of means

Biotechnology vs Telecommunications								
Varables	Obs	t	Sig. (2-tailed)	Mean difference	Std. Err. Difference	Std. Dev. Difference	95% Conf. Interval of difference	
							Lower	Upper
Knowledge Coeherence	21	27.258	0.000	0.085	0.003	0.014	0.078	0.091
Cognitive distance	21	10.462	0.000	0.002	0.000	0.001	0.001	0.002
Gini index	21	15.176	0.000	0.213	0.014	0.064	0.184	0.243
RTV	21	2.573	0.018	0.450	0.175	0.802	0.085	0.815
UTV	21	38.307	0.000	1.056	0.028	0.126	0.998	1.113
TV	21	8.716	0.000	1.506	0.173	0.792	1.146	1.866

Biotechnology vs Electronics								
Variables	Obs	t	Sig. (2-tailed)	Mean difference	Std. Err. Difference	Std. Dev. Difference	95% Conf. Interval of difference	
							Lower	Upper
Knowledge Coeherence	21	27.726	0.000	0.098	0.004	0.016	0.091	0.106
Cognitive distance	21	12.636	0.000	0.002	0.000	0.001	0.002	0.002
Gini index	21	62.894	0.000	0.310	0.005	0.023	0.299	0.320
RTV	21	14.256	0.000	0.717	0.050	0.231	0.612	0.822
UTV	21	-23.139	0.000	-0.507	0.022	0.100	-0.552	-0.461
TV	21	3.614	0.017	0.210	0.058	0.267	0.089	0.332

Electronics vs Telecommunications								
Variables	Obs	t	Sig. (2-tailed)	Mean difference	Std. Err. Difference	Std. Dev. Difference	95% Conf. Interval of difference	
							Lower	Upper
Knowledge Coeherence	21	-19.645	0.000	-0.014	0.001	0.003	-0.015	-0.012
Cognitive distance	21	-22.484	0.000	0.000	0.000	0.000	-0.001	0.000
Gini index	21	-9.162	0.000	-0.096	0.011	0.048	-0.118	-0.074
RTV	21	-1.873	0.075	-0.267	0.142	0.653	-0.564	0.030
UTV	21	37.109	0.000	1.562	0.042	0.193	1.474	1.650
TV	21	10.04	0.000	1.296	0.129	0.591	1.026	1.565

Figure 1 – Count of technological classes (5-years moving average), by sector

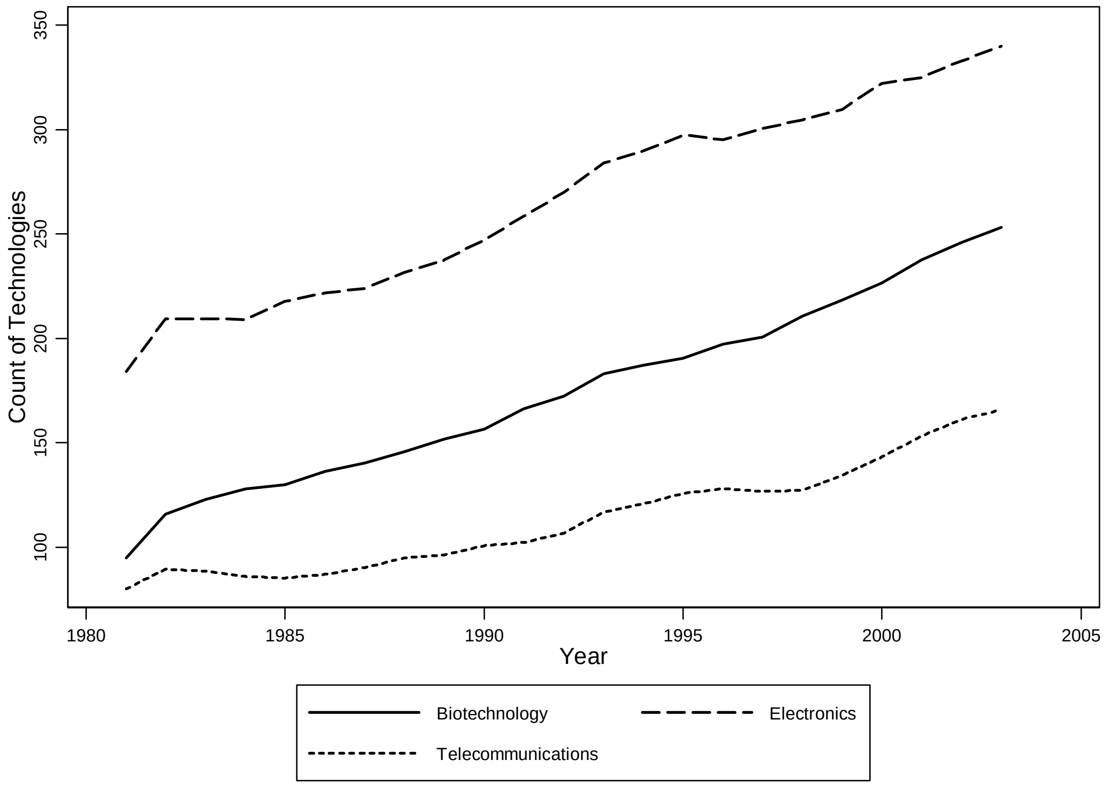


Figure 2 – Evolution of Gini concentration index for co-occurrences of technological classes

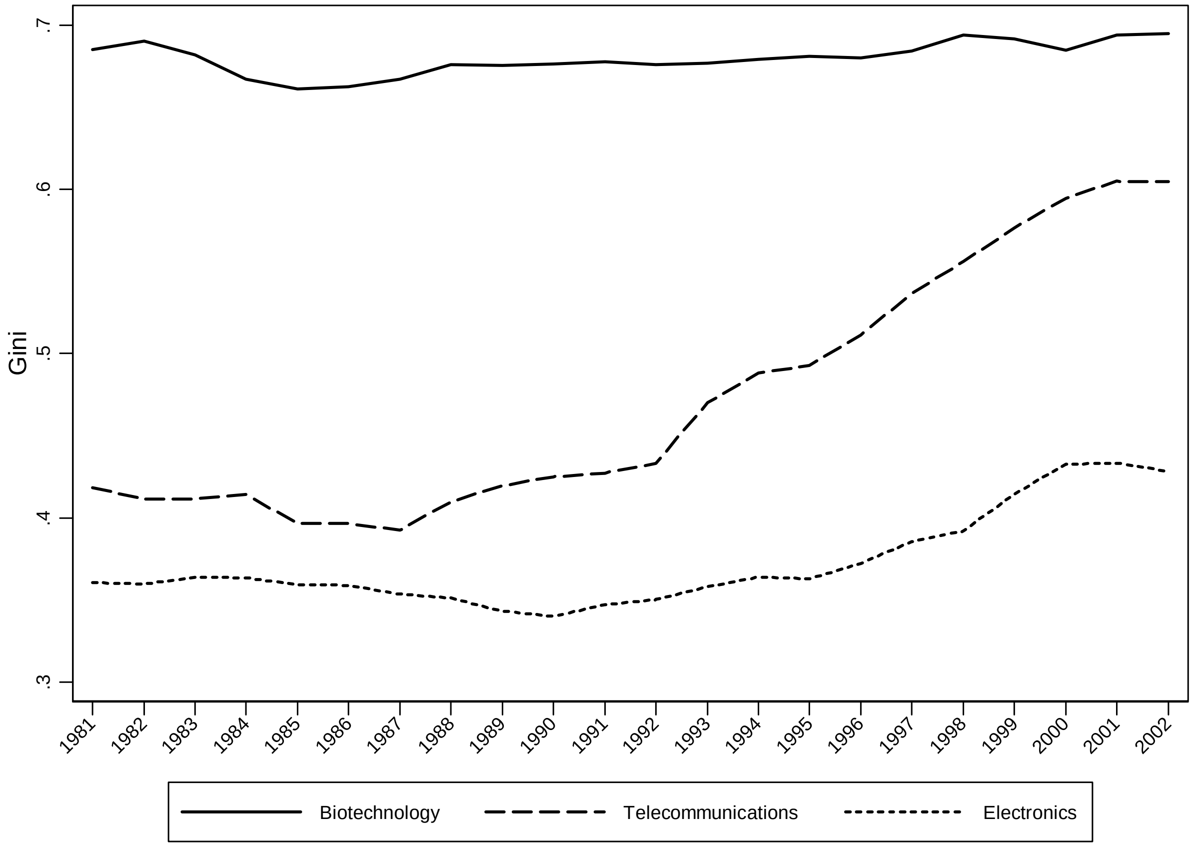
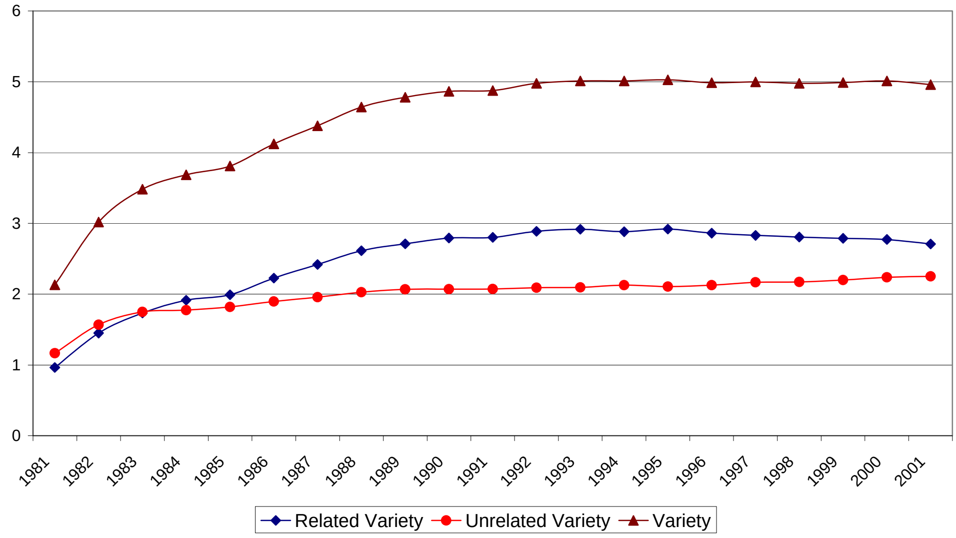
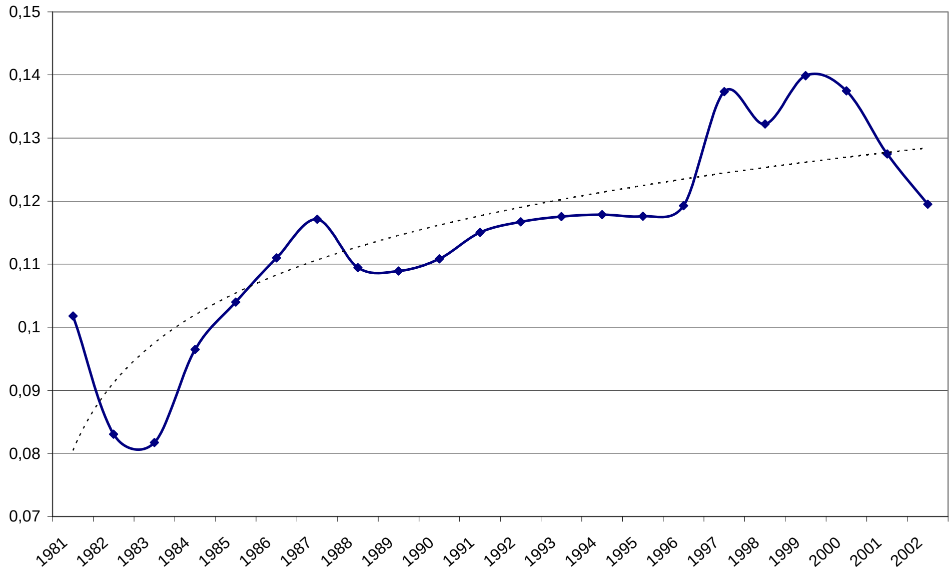


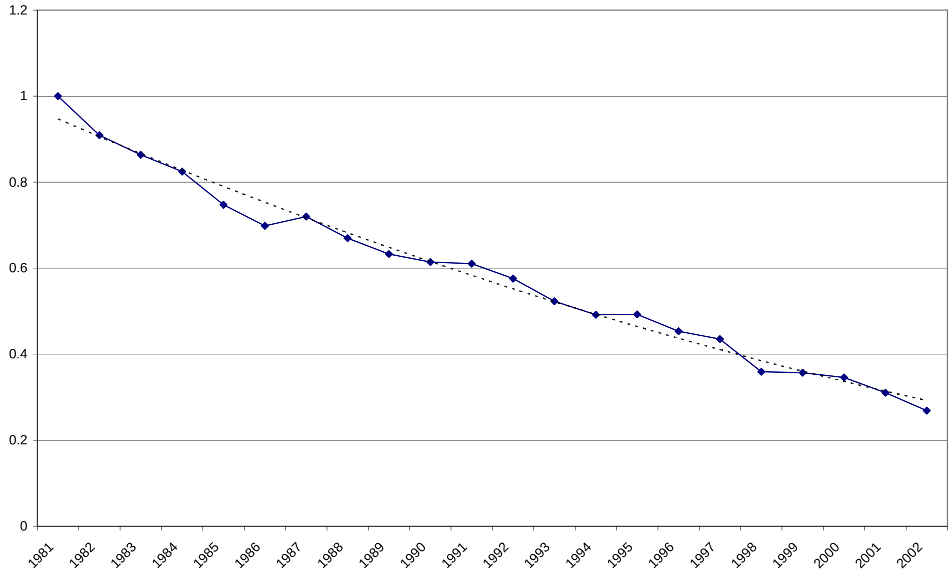
Figure 3 - Properties of Knowledge Base, Biotechnology



a) Variety

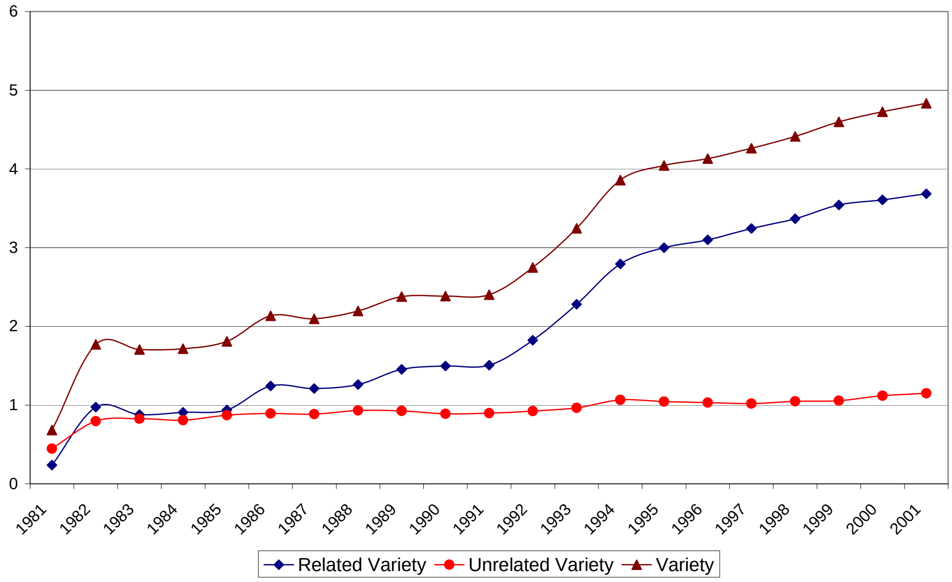


b) Coherence

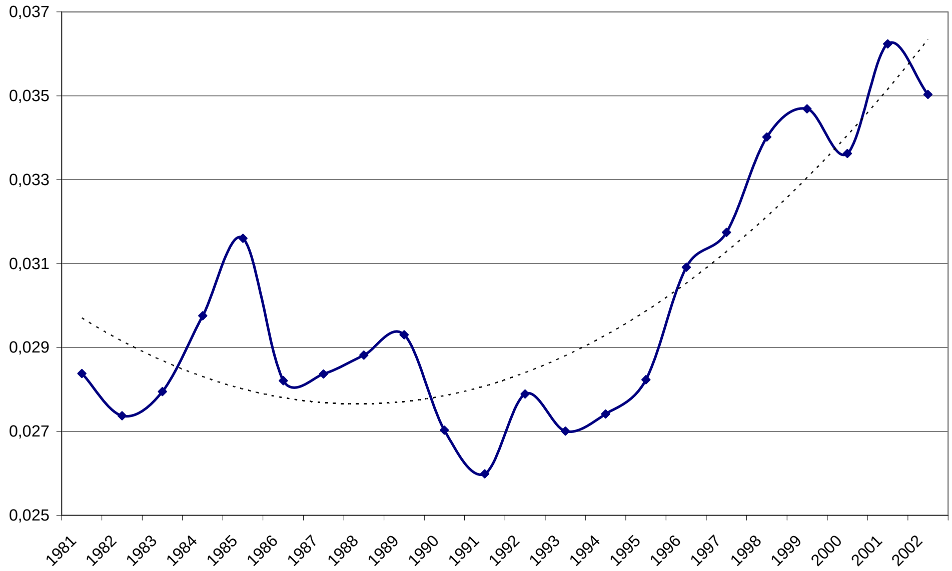


c) Cognitive Distance

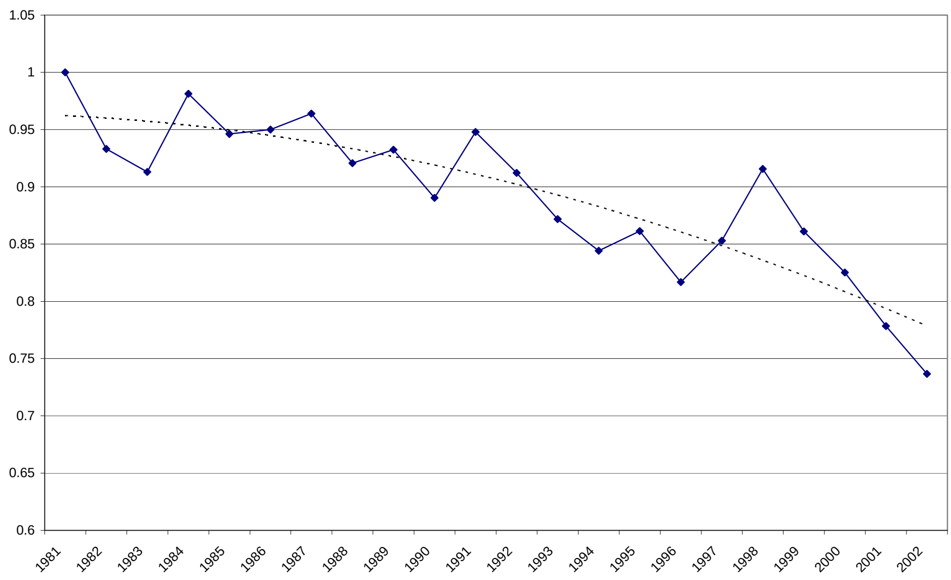
Figure 4 - Properties of Knowledge Base, Telecoms



a) Variety

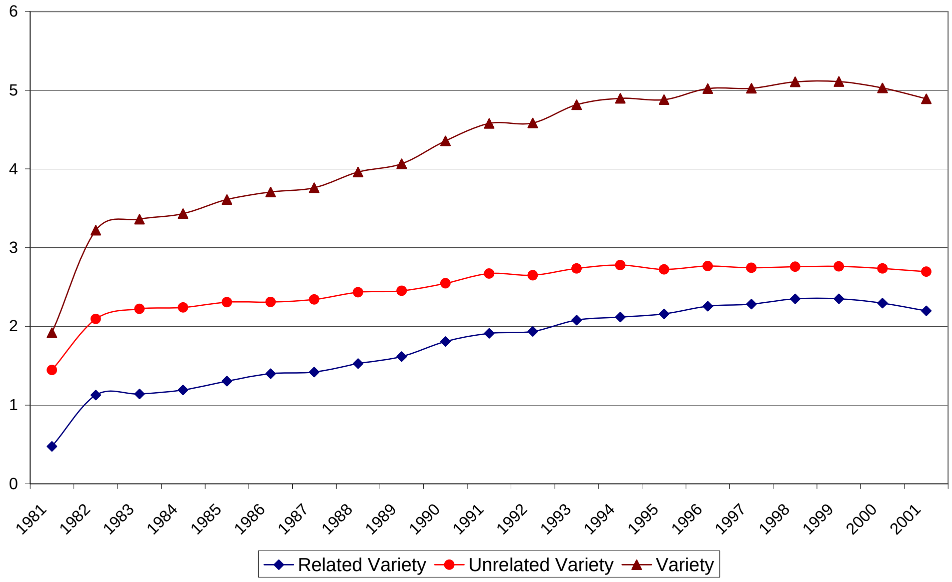


b) Coherence

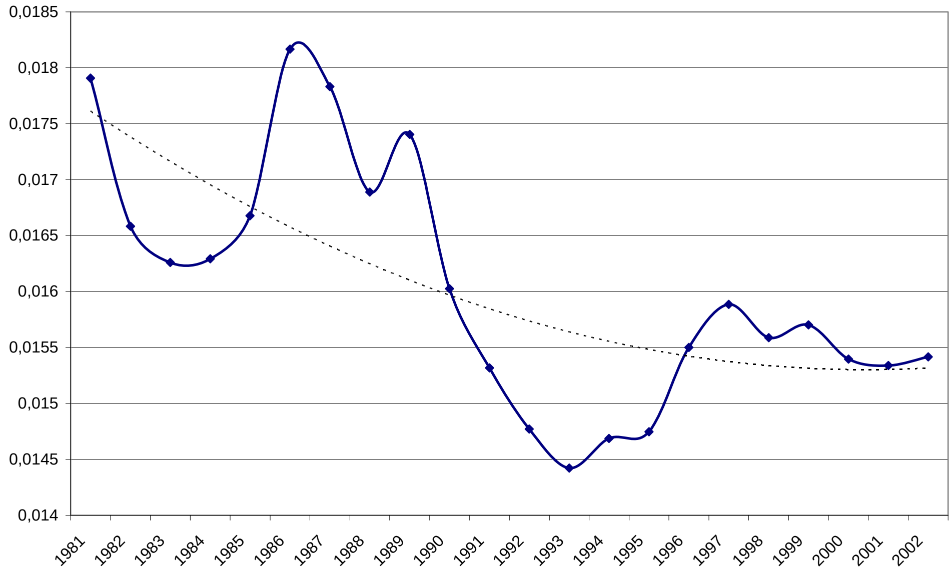


c) Cognitive Distance

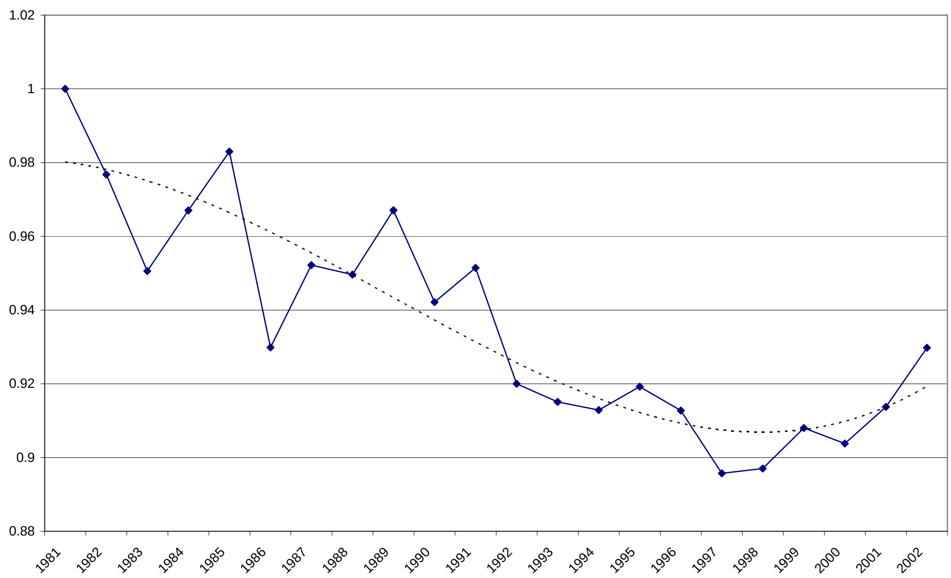
Figure 5 - Properties of Knowledge Base, Electronics



a) Variety



b) Coherence



c) Cognitive Distance

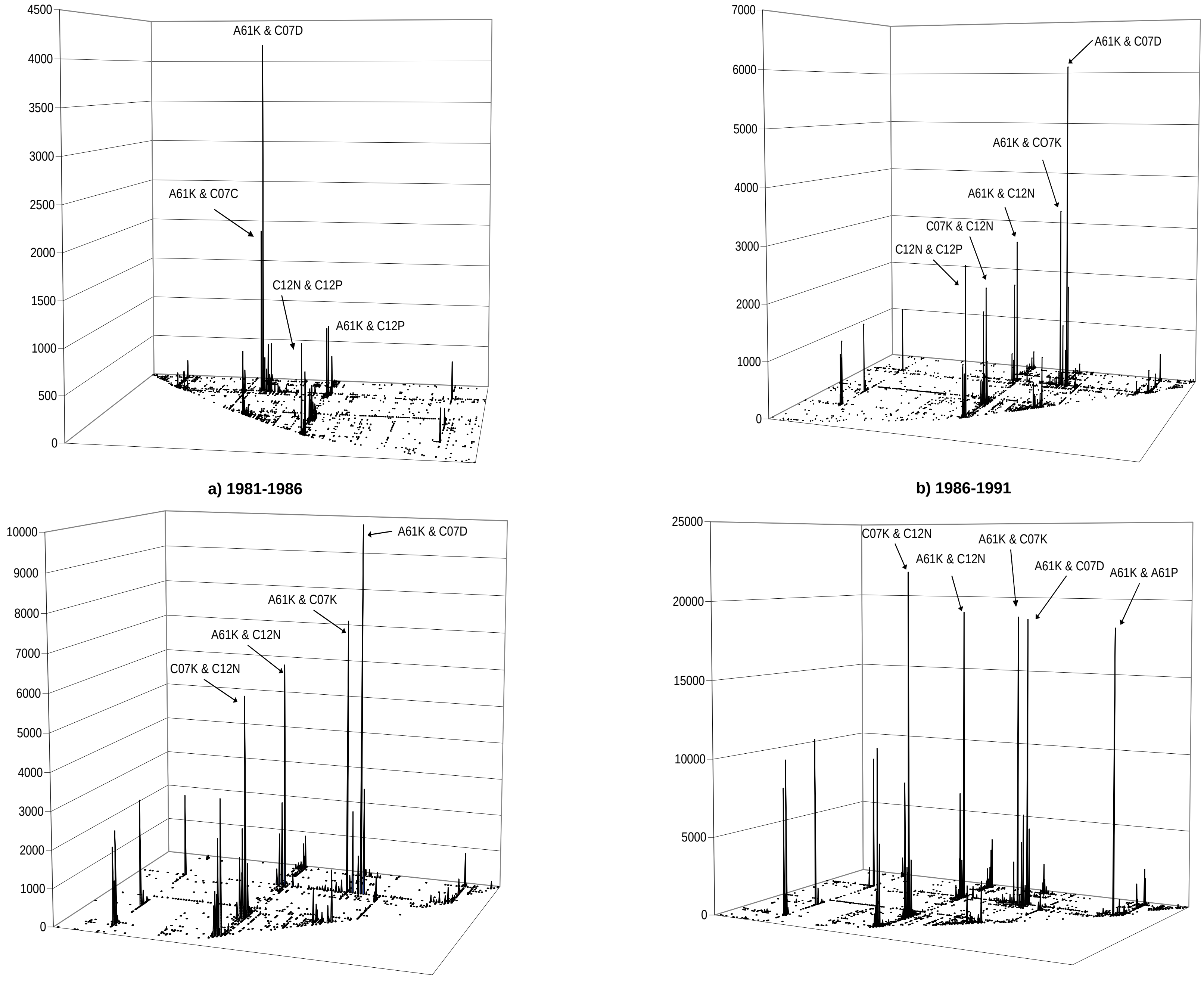
Appendix

Table A1 - Definition of sectors using IPC classes

BIOTECHNOLOGY	
A01H	new plants or processes for obtaining them; plant reproduction by tissue culture techniques
A61K	preparations for medical, dental, or toilet purposes
C02F	treatment of water, waste water, sewage, or sludge
C07G	compounds of unknown constitution
C07K	peptides
C12M	apparatus for enzymology or microbiology
C12N	micro-organisms or enzymes; compositions thereof
C12P	fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture
C12Q	measuring or testing processes involving enzymes or micro-organisms; compositions or test papers thererof; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes
C12S	processes using enzymes or micro-organisms to liberate, separate or purify a pre-existing compound or; processes using enzymes or micro-organisms to treat textiles or to clean solid surfaces of materials
G01N	investigating or analysing materials by determining their chemical or physical properties
TELECOMMUNICATIONS	
G08C	transmission systems for measured values, control or similar signals
H01P	waveguides; resonators, lines, or other devices of the waveguide type
H01Q	aerials
H03B	generation of oscillations, directly or by frequency-changing, by circuits employing active elements which operate in a non-switching manner; generation of noise by such circuits
H03C	modulation
H03D	demodulation or transference of modulation from one carrier to another
H03H	impedance networks, e.g. resonant circuits; resonators
H03K	pulse technique
H03L	automatic control, starting, synchronisation, or stabilisation of generators of electronic oscillations or pulses
H03M	coding, decoding or code conversion, in general
H04B	transmission
H04H	broadcast communication
H04J	multiplex communication
H04K	secret communication; jamming of communication
H04L	transmission of digital information, e.g. telegraphic communication
H04Q	selecting
ELECTRONICS	
F21H	incandescent mantles; other incandescent bodies heated by combustion
F21K	light sources not otherwise provided for
F21L	lighting devices or systems thereof, being portable or specially adapted for transportation
F21M	transferred to F21s and F21V
F21P	transferred to F21s and F21V
F21Q	transferred to F21s and F21V
F21S	non-portable lighting devices or systems thereof
F21V	functional features or details of lighting devices or systems thereof; structural combinations of lighting devices with other articles, not otherwise provided for

G05F	systems for regulating electric or magnetic variables
H01B	cables; conductors; insulators; selection of materials for their conductive, insulating, or dielectric properties
H01C	resistors
H01F	magnets; inductances; transformers; selection of materials for their magnetic properties
H01H	electric switches; relays; selectors; emergency protective devices
H01J	electric discharge tubes or discharge lamps
H01K	electric incandescent lamps
H01M	processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy
H01R	electrically-conductive connections; structural associations of a plurality of mutually-insulated electrical connecting elements; coupling devices; current collectors
H01T	spark gaps; overvoltage arresters using spark gaps; sparking plugs; corona devices; generating ions to be introduced into non-enclosed gases
H02B	boards, substations, or switching arrangements for the supply or distribution of electric power
H02G	installation of electric cables or lines, or of combined optical and electric cables or lines
H02H	emergency protective circuit arrangements
H02J	circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy
H02K	dynamo-electric machines
H02M	apparatus for conversion between ac and ac, between ac and dc, or between dc and dc, and for use with mains or similar power supply systems; conversion of dc or ac input power into surge output power; control or regulation thereof
H02P	control or regulation of electric motors, generators, or dynamo-electric converters; controlling transformers, reactors or choke coils
H04M	telephonic communication
H05B	electric heating; electric lighting not otherwise provided for
H05C	electric circuits or apparatus specially designed for use in equipment for killing, stunning, enclosing or guiding living beings
H05F	static electricity; naturally-occurring electricity
H05K	printed circuits; casings or constructional details of electric apparatus; manufacture of assemblages of electrical components
Source: World Intellectual Property Organization.	

Figure A1 - Matrix of co-occurrences, Biotechnology, 1981-2001



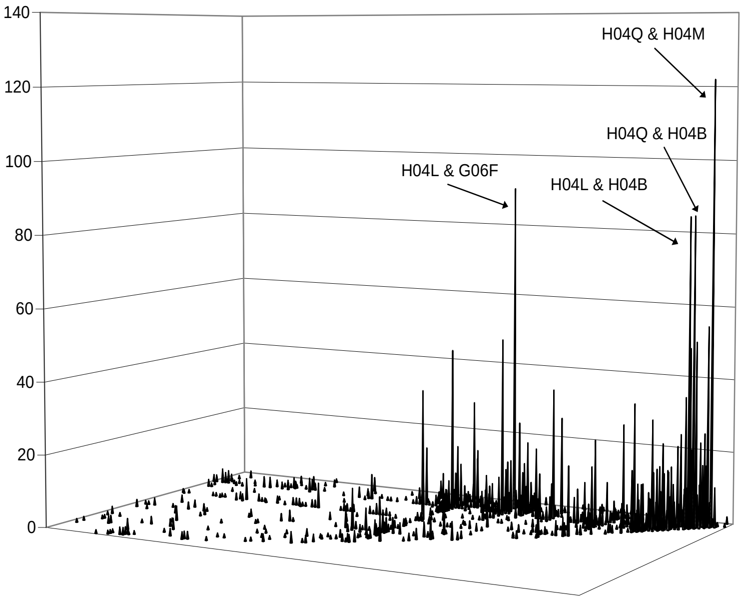
a) 1981-1986

b) 1986-1991

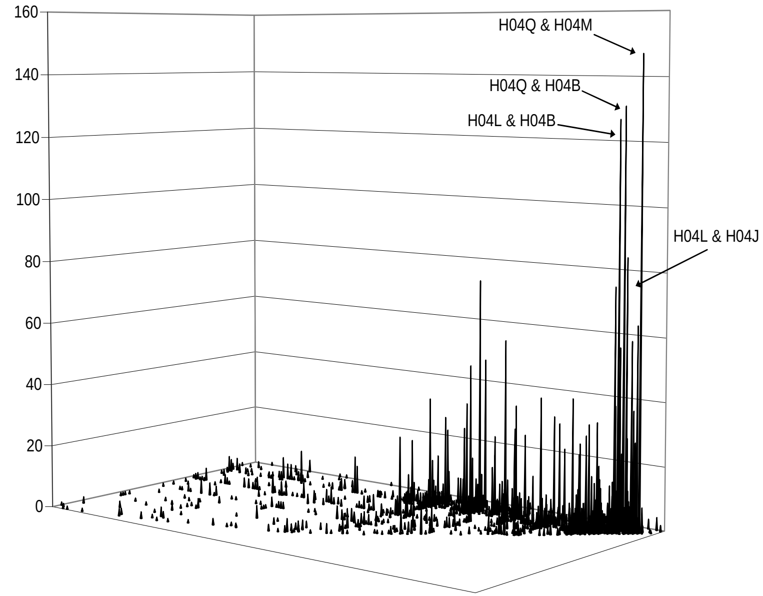
c) 1991-1996

d) 1996-2001

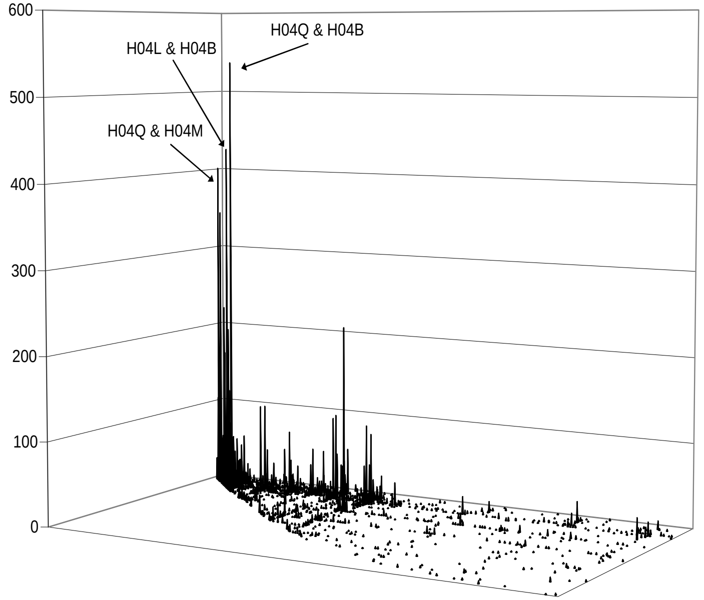
Figure A2 - Matrix of co-occurrences, Telecoms, 1981-2001



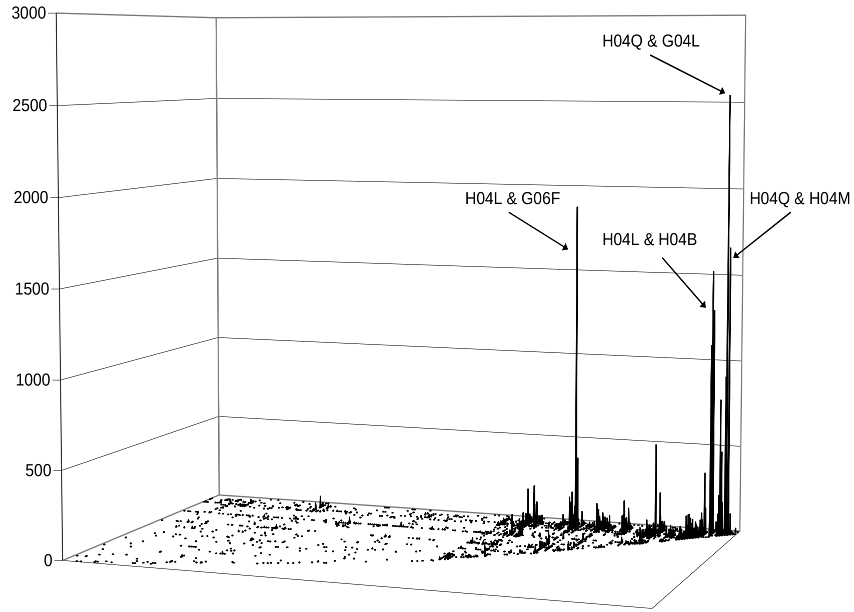
a) 1981-1986



b) 1986-1991

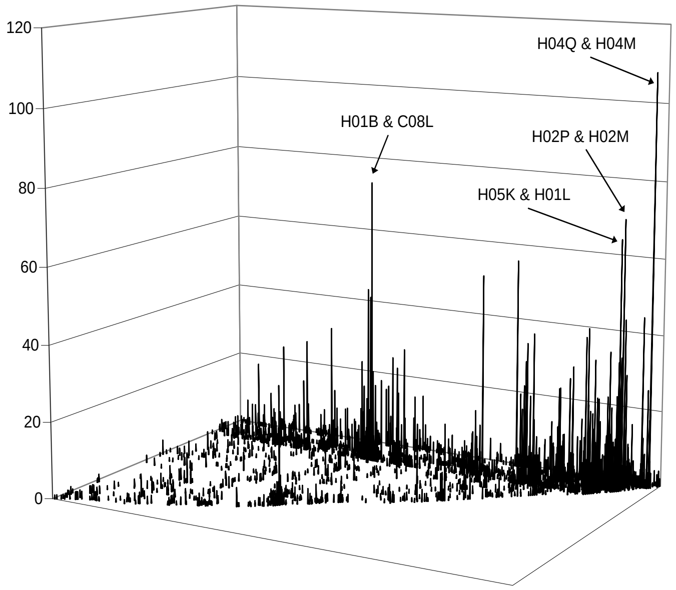


c) 1991-1996

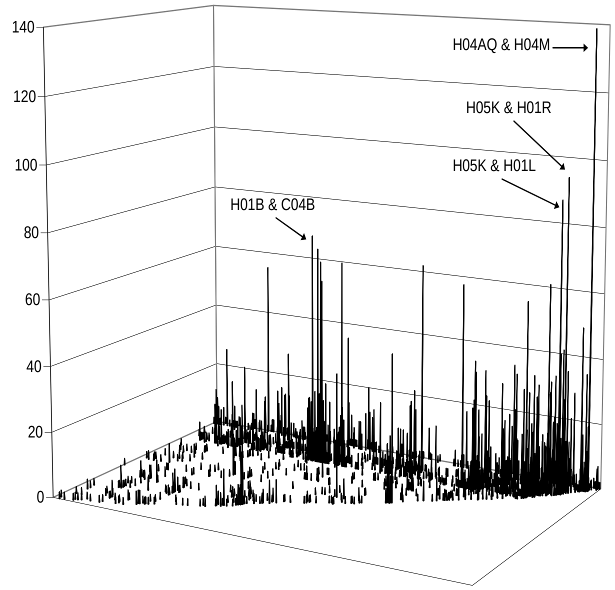


d) 1996-2001

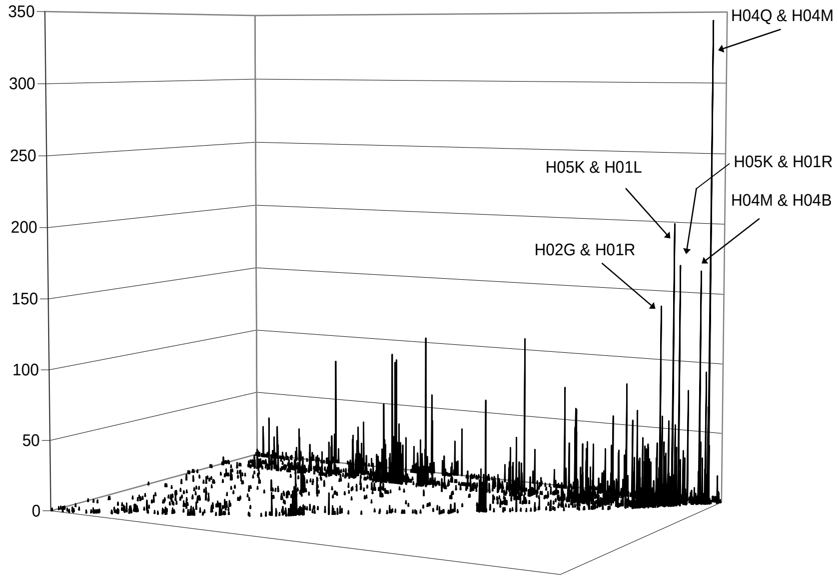
Figure A3 - Matrix of co-occurrences, Electronics, 1981-2001



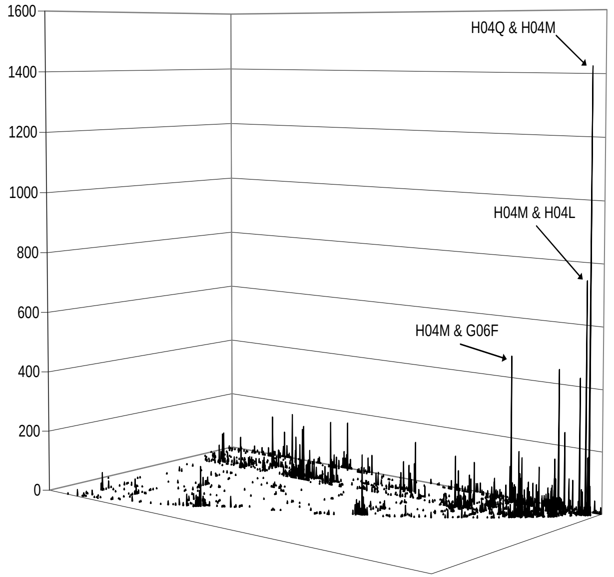
a) 1981-1986



b) 1986-1991



c) 1991-1996



d) 1996-2001