N. Corrocher, F. Malerba, F. Montobbio

The emergence of new technologies in the ict field: main actors, geographical distribution and knowledge sources

2003/37

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

Printed in Italy in December 2003
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

THE EMERGENCE OF NEW TECHNOLOGIES IN THE ICT FIELD: MAIN ACTORS, GEOGRAPHICAL DISTRIBUTION AND KNOWLEDGE SOURCES

Nicoletta Corrocher*, Franco Malerba*, Fabio Montobbio[+]*

*CESPRI - Bocconi University
Via Sarfatti 25, 20136 Milan, Italy
nicoletta.corrocher@uni-bocconi.it
franco.malerba@uni-bocconi.it

[+]Department of Economics
University of Insubria, Varese
Via Ravasi 2, 21100 Varese (Italy)
fmontobbio@eco.uninsubria.it

This paper is based on research done within the TENIA project (The emergence of new industrial activities: fusing manufacturing and services - contract SERD-2000-00194).

We thank Ed Steinmueller for precious advice, Gianluca Tarasconi and Andrea Pozzi for valuable research assistance.

Abstract: This paper examines the emergence of technologies, applications and platforms in the area of information and communication technologies (ICT), using patent data. It detects new technologies/applications/products using *patents' abstracts* and describes them looking at their degree of "hybridisation", in terms of technological domains and knowledge base, at the role of firms in driving the innovation activity, and at the geographical distribution of the innovation. The results show that emerging technologies in ICT are more concentrated across technological classes and across firms than non emerging ones, and that this pattern is invariant across major countries. Furthermore, a preliminary analysis on patent citations show that in emerging technologies knowledge sources are more specific in terms of technological classes and more dispersed in terms of cited institutions. Also there is evidence of a role for universities and public research centres as sources of knowledge.

1. Introduction

The ICT sector represents a key industry in the economy and a crucial source of technical change. Its growing importance is reflected in the increasing number of patent applications as well as in its rising share in total patents. A recent bulletin by Eurostat (2003) shows that, in 2001, the share of the Information and Communication Technology (ICT) sector in the total number of patent applications to the European Patent Office (EPO) was 2.3 times larger than that of 1991. This ratio was 1.3 and 2.0 times larger for Japan and the United States respectively. ICT patent applications to the EPO accounted for 15.5% of the total for the EU in 2001, 18.7% for Japan and 24.6% for the United States. In terms of annual average growth rates, ratios for applications in the ICT sector are well above those of total patents not only for the EU (23.4% vs. 11.0%), but also for the United States (22.0% vs. 10.9%) and Japan (17.7% vs. 11.9%). For patents data the study uses the International Patent Classification (IPC) classes and, following the indication by the OECD, defines as ICT the following classes: Computing, Calculating, Counting (G06); Basic electric circuitry (H03); Electric communication technique (H04).

This paper claims that the use IPC classes to describe technological fields in ICT (and to single out the emerging ones) is subject to some major drawbacks. First, technological progress in this sector proceeds at a very high speed and in many different directions, making it difficult to encompass all the innovations within the existing technological classes. Second, the general purpose character of ICT and the combination/fusion of ICT with other technologies, which has recently been responsible for the emergence of innovations in different sector, make it difficult to assign a patent to a specific class, which is exogenously defined.

This paper examines the emergence of new technologies, applications and platforms in the area of ICT, using patents' data. The main objectives of the research are:

1) to use patents abstract for the identification of relevant ICT-related technologies, products, applications;
2) to show that relevant technologies/applications/products spread across a many IPC classes (also at a 4 digit level of disaggregation). Moreover the set of relevant IPC classes is wider that the one commonly used (e.g. in Eurostat, 2003)
3) to identify among the relevant technologies/applications/products, the emerging ones;
4) to compare the characteristics of the emerging and non-emerging technologies, by looking at the degree of technological hybridisation of the technologies, at the sources of the innovation, in terms of firms, research centres, universities, at the geographical distribution of the innovation. In particular we are interested in the role of big firms and concentration at the firm level to promote or hinder the emergence of new technologies/applications/products in ICT.

The paper is structured as follows. In the first part (Section 2), we discuss the general issue of the emergence of new technologies and industrial activities, introducing the concept of hybridisation. In Section 3 we describe the methodology used to identify

ICT-related technologies. The methodology involves the selection of relevant *triples* of words from patents' abstracts, which allow to detect, without a subjective bias, existing and important applications in the ICT area, which extend over different technological classes. The methodology draws its theoretical background from the existing works on keyword and co-word analysis (Courtial, Callon and Sigogneau, 1993; Van Raan and Tijssen, 1993; Noyons and van Raan, 1998; Ding et al., 2000) and aims at overcoming the existing classification system (IPC), which does not properly capture the continuous and complex technical progress in the area and the general purpose nature of some ICT-based platforms and applications.

In Section 4 we describe the set of ICT-related *triples*, by distinguishing and categorising different types of technologies/applications/products. Moreover we show their characteristics in terms of distribution across IPC classes, distribution across firms and countries. In Section 5 the emerging technologies/applications/products are identified and we compare the characteristics of emerging and non emerging technologies. The results show that emerging technologies in ICT are more concentrated across technological classes and across firms than non emerging ones, and that this pattern is invariant across major countries. Furthermore, a preliminary analysis on patent citations show that knowledge sources are more dispersed in emerging technologies as compared to non emerging ones and that there is a role for universities and public research centres as sources of knowledge.

2. The emergence of new technologies: some general considerations

This paper focuses on the nature of the emerging technologies in terms of the characteristics of their knowledge sources and of the actors who bring them about. As far as knowledge sources are concerned, the main issue is related to whether new technologies stem from a single idea within a selected and homogeneous set of technological principles, or instead are the result of the convergence of different ideas from different technological fields. Furthermore, we are interested in understanding which of these technologies serve as a source of innovation in different areas. Regarding the actors involved in the innovation process, three issues are addressed in this paper. First we enquire the role of industrial concentration of innovative activities in the promotion of new technologies. Secondly, we analyse whether there is a distinct pattern in the creation of new technologies at a country level. Finally, we investigate the role of different institutions - universities and public research centres - in the development of new technologies.

The work of Pavitt has emphasised that emergent technological paradigms spread across different industries (such as the digital technology one). Moreover he has underlined that increased technological complexity - arising within the high tech sectors - generates innovations that are more and more developed across industrial boundaries (Koumpis and Pavitt, 1999; Mahdi and Pavitt, 1997). In particular, there is a growing tendency towards the fusion of existing technologies and towards the emergence of applications that spread across different technological areas (Kodama, 1992; Miyazaki, 1994; Koumpis and Pavitt, 1999; Fujimoto et al., 2000). This is particularly true for the so-called general purpose technologies (Bresnahan and Trajtenberg, 1995; Helpman, 1998), which are characterised by pervasiveness of use and inherent potential for technological improvement and dynamism. In this sense, technologies such as ICT play the role of enabling technologies, opening up new opportunities in different fields.

A useful starting point is to consider that new technologies and applications undergo life cycles processes. They emerge out of the knowledge and capabilities related to existing technologies and are initially aligned to the problems of the old regimes. In the first stages, the domain of application of the new technology is quite limited and public demand often provides the initial niche market for it, since in that area performance is more important than costs and considerable financial resources may be available (Rip, 1995). This has been evident in the case of the development of the digital computer regime and of the Internet, whereby the requirements of the US Defence Department have strongly stimulated technological advances in the area. The first commercial domains are crucial for the take off of a new technology: they initiate learning processes at the supply and user side and foster institutional support from investors, customers and other actors. At the very beginning, the existing technologies may benefit from dynamic learning effects related to the new technologies, and their trajectories are often sustained by the interests of different parts in their continuation - e.g. incumbent firms that have invested in infrastructure based upon one specific technology (Ehrnberg and Jacobsson, 1997). However, once the new technologies become more robust and are accepted in the market, they start benefiting from dynamic scale and learning economies, from the development of complementary innovations and from institutional adaptations, so that irreversibilities emerge (Van den Ende and Kemp, 1999).

The emergence of a new technology is conceptualised in this paper as an evolutionary process of technical, institutional and social change, which occurs simultaneously at three levels: the level of individual firms or research laboratories, the level of social and institutional context, and the level of the nature and evolution of knowledge and the related technological regime.

2.1 Knowledge Sources and Technological Hybridisation

A lot of new technologies stem out from established technologies, as a combination of existing knowledge. As Tushman and Anderson (1986) stress, technological change is a cumulative process punctuated by a major advance. Within this framework, it is possible to distinguish between technological discontinuities that generate new product classes (airlines, automobiles, computers), by product substitution (transistors vs. vacuum tubes, diesel vs. steam locomotives, steam ships vs. sailing ships), or by fundamental product improvements (jets vs. turbojets), and technological discontinuities that generate process substitution (e.g. thermal vs. catalytic cracking in crude oil refining), or process innovations that result in radical improvements in sector-specific dimensions. However the co-ordination, combination and integration of existing technological competencies and knowledge is an important characteristics of emerging technologies and may take place both within different manufacturing industries, as well as between manufacturing, new materials and services, as the cases of mechatronics, computational chemistry, multimedia show (Kodama, 1992; Mahdi and Pavitt, 1997; Mansella and Steinmueller, 2000).

The clustering and combination of different technologies - whether of existing technologies, or of existing and new ones - occur in many technological systems, in which different interrelated subsystems have to work close to each other to make the entire system efficient. Telecommunications are an interesting case, whereby different technological subsystems (terminal equipment, local access, switching, long-distance transmission, signalling and control) interact and function as a technically integrated end-to-end system that provides the user with a range of voice, data and image services (Davies, 1996). In the course of the twentieth century new scientific areas such as electronics and genetics have emerged, new technologies have been introduced by firms, universities and research centres, and new products and services have been demanded by consumers. New opportunities now spread across a wide range of technological fields while the existing technologies are more and more employed as complementary inputs in the development of new processes and products.

According to a general perspective (Kodama, 1992), this phenomenon identifies the combination of incremental technical improvements from previously separate technological areas, so that new products and/or activities emerge, bringing about radical changes in existing markets and possibly creating new ones. It is possible to envisage fusion as a process which stems from the convergence of similar technologies or of related technologies, or even of unrelated technologies that become part of a common paradigm (e.g. computing and communications technologies; ICT and audio-visual technologies). Within this context, the fusion may occur at different levels, since the technologies into question may simply be coordinated, or may be combined, or may even be integrated. Through these means of interaction, the fusion of technologies drives a transformation of the existing activities into new *hybrid activities*, which span across different technological fields and even across different industries.

ICT is of the most important general purpose technologies that has generated technology fusion and has allowed the emergence of an increasing number of new technologies and applications, which are used in different field. One could ideally envisage the result of the fusion of manufacturing and service knowledge in the activities related to the ICT as a *platform*, which represents a combination of hardware, software and specific knowledge of the service end user market. In this case, the interaction has as a main result the fact that specialised applications provide the basis for the development of specific hardware platforms. At the same time, technology plays an important role in the evolution of information-intensive service providers, since these firms rely upon technological platforms to deliver their services (Pistorious and Utterback, 1997). In the context of multimedia activities, for example, the combination of audio-video hardware and software with the competencies of the entertainment industry has allowed the replacement of analogue production systems that use films as the medium, with the digital systems, that use computer generated images as the medium. Furthermore, the growth of multimedia and new digital media in general has significant implications for the emergence of new types of equipment and local infrastructure, in that the network structure will ultimately be determined by the availability of information services (Mansell and Steinmueller, 2000).

This discussion brings about the following conclusions. First, we can expect that, along the evolution of a technology, the range of its applications expands over time across different technological fields. This implies that, in their emerging phase, technologies are embedded and exploited in applications and inventions that are relatively more concentrated in specific technological classes. Second, we can expect to find a relatively higher amount of hybrid activities in established technologies as compared to new technologies. Finally, we do not have any strong a priori assumption on the degree of dispersion of the knowledge sources of invention and applications in new technologies, since this depends on the nature of the innovation. Radical innovations may rely on a differentiated knowledge base, while incremental innovations may constitute an improvement along a very specific technological trajectory.

2.2 Countries, Structure of Innovative Activity and Institutions

The processes outlined above involve linkages between different actors - firms and users, universities and research organisations, institutions. These linkages that are often country and sector specific affect technical change, which in turn can radically modify the structure of countries, firms and institutions within which it emerges. Co-evolution exists both between different technologies (Pistorious and Utterback, 1997) and between the technology and the surrounding actors: firms, non-firm organisations and institutions, demand, social and sectoral environment (Nelson, 1994; Metcalfe, 1998; Malerba, 2002). Here we consider the process of creation of applications and inventions along three dimensions: countries, market structures and institutions.

Countries

Along the technological life cycles, the product cycle hypothesis of the locational evolution of innovative activities (Vernon,1966 and 1974; Grossman and Helpman, 1991; Krugman, 1979 and 1995) suggests, in its baseline version, that products are manufactured and consumed primarily in the most advanced countries because of economies of scales, imperfect competition, high income elasticity of demand, high labour costs, first mover advantages, and because there is a need for a swift communication between the market and the firm. The locus of production might shift in a subsequent phase as products become standardised. According to this view, we should expect to find the applications of emerging technologies relatively concentrated across countries, while inventions related to established technologies should be more dispersed, because of de-localisation of multinational corporations and because of standardisation, which drives innovation and imitation by new firms in less developed countries.

Structure of Innovative Activities

In terms of innovative activities, it is worth underlining that firms have become multi-technological in nature: the development of artefacts and production processes no longer draws upon just a few technologies, but upon a broad range of technologies. The literature has shown that within a specific sector more than one technology may be relevant, so that in principle, it is possible to build a technology-product matrix that links the products to several technologies (Kodama, 1992; Granstrand et al., 1997; Fai and Von Tunzelmann, 2001). Thus firms in an industry may be active in several technological fields that do not traditionally identify its core activities. Fai and Von Tunzelmann, (2001) show, for example, that the chemical industry is technologically strong in the mechanical and transport fields related to the equipment, as well as the mechanical industry is quite strong in the electronic and in the chemical fields. According to this view, firms move out of technological fields which were important in the past and move into other areas that are relevant in the present or in the future. In particular, the authors show that as the core technologies of firms mature, they would mainly look downstream to user applications as a first way of diversifying. In the case of digital electronics technology, the phenomenon of technological convergence between sectors implies that eventually companies belonging to a sector need to integrate the new technologies within their own activity and to achieve a certain degree of technological diversification (Granstrand et al., 1997). The increasing technological diversification within firms, coupled with the increasing technological diversity across firms, stems from the continuous emergence of opportunities related to new technologies and is reflected in the changing technological structure at the industry level.

This line of enquiry points at the important role of big established firms in promoting new applications and inventions (Pavitt, 1994; Patel and Pavitt, 1997). This relates to the analysis of the relationship between the structure of innovative activities and the rate of technological change and to the old debate on market structure and innovation. In general, that debate has come to the conclusion that, up to a certain level of concentration (or up to a low entry rate), higher concentration (lower entry) increases innovation. Concentration increases the incentives to innovate, because of the availability of internal financial resources necessary to invest and because of cumulativeness of learning and technical advance (Dasgupta and Stiglitz, 1980; Nelson and Winter, 1982; Malerba, Montobbio, 2003).

However the technological life cycle model suggests that the initial phase of the product cycle model involves a systemic or radical innovation (Klepper, 1996). This means that a new design concept emerges and a social or technical function is performed in a completely new way. In these circumstances, organisational change may be a difficult and slow process, and the ability of large organisations to respond quickly to competence-destroying change can be quite limited, since they are endowed with a specific and historically rooted set of skills (Patel and Pavitt, 1997). Therefore, radical innovations may be associated with firms that are new, or small- and medium-size. New firms are likely to develop the appropriate structure for a new and uncertain environment, especially if they are characterised by high skill intensity, limited hierarchy levels and frequent redefinition of tasks (Utterback, 1994).

Concentration appears in the market when a dominant design has already emerged and the product is more standardised. A few firms enter the market, as long as the rate of entry is related with new products, and barriers to entry are high. When the product is standardised, incumbents have a competitive advantage based on their established routines. As a result, the phenomenon called "industry shake-out" is likely to occur and is originated by a decreasing rate of entry and by the exit of the less efficient firms, which are not able to supply a competitive dominant design (Klepper, 1996 and1997)[1].

[1] Gort and Klepper (1982), and Klepper and Graddy (1990), working on a longitudinal data base for 46 new products, find that the number of firms initially grows, in a second stage declines, and finally levels

We do not have strong a priori assumptions on the concentration of the creation of inventions and new industrial activities across firms. This can occur within very large, multi-technology firms through a process of diversification and innovation, as well as within very small firms exploiting new opportunities emerging from research in other firms, universities and public research centres.

Universities and Public Research

New technological and scientific knowledge may also arise from research performed by public institutes and universities. In particular, we believe that the research output by public institutions and universities may be an important guiding and enhancing factor for emerging technologies. This is especially true when public research has tangible outcomes in developing the knowledge bases in specific core technologies (Nelson, 1993). However science, and therefore universities and public research centres is a major source of new technologies, but science is also affected by and follows the technological developments (Rosenberg, 1976). Many sectors (like biotechnology) took off in the US from a small group of academic discoverers and the industry soon clustered regionally around leading-edge research centres. The role of "star scientists" is not to be solely related to the early stages of the industry development. The presence of a critical mass of excellence in science appears to have guided localisation of science based industries firms, even after basic knowledge diffused and techniques got routinised. Empirical studies provide considerable evidence suggesting that the timing and location of new biotechnology firms is primarily explained by the presence of scientists who are actively contributing to the basic science (Zucker et al., 1994).

In the empirical section of this paper, we analyse whether universities and public research centres play a significant role in ICT-related technologies. This is done investigating the creation of the upstream knowledge base. In this context, we also analyse the differences between emerging and non emerging technologies.

3. Methodology

We examine the emergence of new technologies, products and technological applications related to ICT by adopting an innovative approach, which is based upon the identification of relevant triples of words from patents' abstracts. The literature on keywords has shown that there are two main ways to extract words from a journal article, a paper, and abstracts. The first one consists of extracting words from keyword lists, titles, and sometimes even classification codes. Coulter et al. (1998) for example select descriptors chosen by professional indexers, considering that their experience guarantees a correct procedure of keyword selection. Similarly, Noyons and van Raan (1998) utilise the co-occurrence of classification codes. This methodology has a main shortcoming, in that indexing might reflect the preconceptions and points of view developed by indexers during the course of their training and the probable inconsistencies in keyword selection by professional indexers working for different databases. The second way of finding words involves extracting words directly from

off. At the same time, they find that the length of each stage is highly product-specific. Moreover, new products innovations turn out to be correlated with positive net entry.

full-text documents by using some software or by developing specific algorithms. The words or sets of words with a certain frequency are chosen as the unit of analysis to represent the core topics of the specific field.

Our methodology is similar to the latter described above. It takes patents' abstracts as the unit of analysis, since they provide a comprehensive description of the technology, product or process to be patented, as well as of the potential applications of that technology, and allow to capture the links between different knowledge domains and different technological fields[2]. The procedure entails the extraction of relevant words from patents' abstracts and the subsequent identification of relevant technologies/products/applications.

Our starting point is a extremely wide set of technological classes that might refer to ICT. We select eight technological classes belonging to the sections of *Physics (G)* and *Electricity (H)* of the International Patent Classification (v.7): **G01** (measuring; testing); **G06** (computing; calculating; counting); **G09** (educating; cryptography; display; advertising; seals); **G11** (information storage); **H01** (basic electric elements); **H03** (basic electronic circuitry); **H04** (electric communication technique); **H05** (electric techniques not otherwise provided for). This choice expands the set of technological classes usually chosen when analysing patents in ICT. Since the main objective of the study is to capture and describe emerging technologies, we take into consideration a time period of recent years (1995-1999)[3]. Our final sample consists of 102547 patents' abstracts.

Starting from the sample of patents' abstracts, one possibility would be to identify the occurrence of pre-selected ICT-related keywords in patents. However, as underlined by the literature and as emerged in some preliminary analysis[4], such an approach has a major limitation in that, by making use of pre-conceived keywords, it identifies already existing and relevant technologies, products, and platforms, implying a strong degree of subjectivity in the analysis. While this kind of work can lead to interesting results, it does not provide an immediate picture of the actual technological content of patents' abstracts, and even more, it does not allow to detect all the relevant technologies, especially the emerging ones. In order to make a further step, we couple this traditional keyword approach above mentioned with a "bottom-up approach", with the objective to identify endogenously relevant technologies without imposing any pre-determined keyword when examining the content of patents' abstracts. This bottom-up methodology requires:

1) the extraction of triples of words within patents' abstracts;
2) the selection of triples that appear with a significant frequency;
3) the identification of relevant triples, i.e. of triples which represent a technology, product or technological application.

[2] The analysis relies upon the EPO-CESPRI dataset. Years are assigned to patents using the priority dates (see below section 4.2).
[3] It is worth mentioning that there is a lag of 18 months between the application date of a specific patent at the European Patent Office and the inclusion of that patent in our database, which means that data from 2000 and 2001 are not used for this analysis.
[4] In a previous work (Corrocher and Montobbio, 2003), we performed a simple keyword analysis, choosing a-priori some relevant technologies, product and applications in the ICT area (e.g. *Internet Protocol*, *Wireless Application Protocol*, *Code Division Multiple Access*, *Multimedia*), which were derived by text-books and specialised journal articles, and examining their occurrence within patents.

Some caveats and limitations of the analysis should be emphasised. The triple extracted depends upon the language which is used in drafting the patent application. This may be subject to two different types of distortion. First there may be a strategic use of the language. Applicants may tend to hide relevant keywords to influence the patent examiner. Second there may be a country specific use of the language (grammatical structures). Of course we cannot control for all this aspects. However we can claim that the use of patents abstracts provide a partial improvement with respect to simple keywords (where the impact of strategic behaviour of applicant may be even stronger) and that patents abstracts are revised during the process leading to publication and therefore we can expect that some language specificities might be corrected.

The present analysis, which is performed on patents' abstracts (instead of patents' titles as done elsewhere, Courtial et al., 1993), involves first the extraction of words from patents' abstracts and requires the development of an ad-hoc algorithm. We assume that a technology or a technological application is adequately identified by a sequence of at least three words, since previous analysis (Corrocher and Montobbio, 2003) aiming at the extraction of couples of words did not provide significant results. The algorithm is therefore developed with the aim of extracting triples of sequential words[5]. From the initial data set of 102547 patents we extracted more the seven million triples. We have then selected the triples that occur in at least 30 patents[6].

It is important to underline that, within the provided list, there are still some generic terms that are very likely to be used in description of any modern electronic system and that provide only limited distinguishing information. For example, *printed circuit boards* and *wiring circuit boards* tell us that this is a multi-component system, and the central or data processing unit and memory tell us that the device has some intelligence. Since virtually all systems of interest are multi-component and have some degree of processing capacity, the terms do not add much. Among the generic keywords, there are some whose relevance is difficult to detect. A good example in this respect is *light emitting diode*: while virtually any modern electronic system may incorporate an light emitting diode (e.g. a power on indicator), the inclusion of this detail in a patent may indicate an essential element of the system (e.g. an infra-red controller or an emitter that is meant to couple with a receiver).

A relevant methodological issue concerns the handling of triples that represent the same product/application, but appear as two separate keywords (e.g. *radio base station* and *base radio station*): these triples need to be grouped, so that they both identify the same

---

[5] The examination of the initial list of triples highlights the existence of some problems. First, there are many meaningless triples that are made of articles, prepositions, adverbs, verbs and so on: in order to solve this problem, an a priori "cleaning" of patents' abstracts is required. In doing so, it is worth underlining that, by deleting some elements in the abstracts, words that were previously separated now become very close to each other. Second, because of punctuation, triples of words that indicate the same technology appear as two different triples (i.e. "communication network," "communication network."). Taking into account these two types of problems, the analysis proceeds with the cleaning of the abstracts and with a second extraction of triples of words, which results in about seven millions triples.

[6] It is important to mention that a triple may occur more than once in the same patent. We have eliminated the duplications so that the frequencies reported here identify the exact number of patents in which each single triple appears.

product/application. In order to group similar keywords, or keywords that clearly refer to the same product/application, we proceed in two steps: first we analyse the complete dataset of triples to detect the most evident similarities, then we perform a co-word analysis of triples[7]. The existence of many co-occurrences around the same word or couples/triples of words identifies a strategic alliance within texts that may correspond to a specific topic. Co-word analysis reveals patterns and trends in a specific discipline or technological field by measuring association strengths of terms. The main advantage of this methodology is that it visualises the intellectual structure of a specific scientific or technological field into knowledge maps of this field.

Here the co-word analysis, differently from the traditional literature, does not have the aim of building the knowledge map of a specific technological field, but allows us to detect triples that represent the same technology or product or application. For example, a high co-occurrence of the triples *code division multiple* and *division multiple access* means that these two triples can be considered as referring to a single technology (*code division multiple access*). The co-word analysis also permits to highlight, at a very preliminary stage, the possible technological and knowledge links between different technologies in the ICT area. However, at present it does not seem to reveal significant patterns between the existing triples and therefore it is useful as a means of corroborating the process of grouping of different keywords.

The final result of this analysis is a list of 119 triples which constitute relevant technologies, applications, platforms or products in the ICT area (for a selection, see table A1 in the appendix). It's worth mentioning that those triples have been endogenously selected and are not based on a preconceived set of keywords. Starting from this dataset, we perform a descriptive analysis aimed at identifying the distinctive features of the selected technologies and, in particular, the characteristics of the emerging ones.

4. A preliminary analysis of ICT-related technologies

4.1 Some general definitions and description of the dataset

This section is devoted at providing some general descriptions on the nature of the technologies we have selected. In this context, we aim at investigating first, whether the IPC constitutes an appropriate system to classify inventions in the ICT domain, and

[7] This part partially draws upon the existing literature on co-word analysis, which counts and analyses the co-occurrence of keywords in the publications of a given subject and has the potential to map the relationship between concepts and ideas in sciences and social sciences. It reduces and projects the data into a specific visual representation with the maintenance of essential information contained in the data (Ding et al., 2000). It is based on the nature of words, which are the important carrier of scientific concepts, ideas and knowledge (van Raan and Tijssen, 1993). Relevant applications of this methodology can be found in different fields such as software engineering (Coulter et al, 1998); scientometrics (Courtial, 1994); neural network research (Noyons and van Raan, 1998; Van Raan and Tijssen, 1993); patents (Courtial et al. 1993); medicine (Rikken et al., 1995). A relevant issue for the present work is that this type of analysis relies upon the assumption that the text's keywords constitute an adequate description of its content. In particular, the set of keywords co-occurring within the same paper are an indication of a link between the topics to which they refer (Cambrosio et al., 1993).

second, if some technological classes are more relevant than others in capturing the technological change in the ICT area.

The empirical analysis is based upon the investigation of the occurrence of these technologies within patents' abstracts. Generally, laws require that, in order to be patentable, the invention must be new, it must involve an inventive step (i.e., it must be non-obvious), and it must be industrially applicable. A patent document contains two types of information: bibliographic information and technical information. The bibliographic information includes, among others, the abstract of the description of the invention outlining the existing technical background and knowledge (the "prior art") on which the invention is based, as well as the contribution the invention makes to solve the technical problem involved. The concept of industrial applicability of the invention implies that the problem must be a technical one[8]. From this it is reasonable to argue that we can use the different triples as relevant technological classes and compare them with the IPC technological classes and that the simple counts of patents which display in their abstract the triple can be used as a measure of innovative activity in that specific technological class.

The relevant point here is to underline that our triples are different from the technological classes of the IPC. The IPC is a hierarchical classification system comprising: Sections; Classes; Subclasses and Groups (main groups and subgroups). The primary aim of the IPC is the establishment of an effective search tool. To this end, it attempts to ensure that any technical subject with which an invention is immediately concerned can be classified, as far as possible, as a whole and not by separate classification of constituent parts. This tendency towards preservation of the complete character of an invention, rather than its dissolution into component aspects, is further reflected in the fact that, to a considerable extent, the IPC employs the principle of classifying inventions according to their intrinsic nature (the "function-oriented" principle), rather than their possible applications, although there are applications places that are intended to cover completely the classification of technical subjects in the disclosure of which the only important information relates to a particular field of use. The function-oriented places embrace a wider concept in which the construction or functional characteristics of a subject are applicable to more than one field of use, or in which the application to a particular field of use is not considered essential.

Our methodology instead is more oriented towards the identification of different technologies/applications/products with different knowledge bases that may belong to different IPC classes. and technologies within the same abstracts and allows the detection of hybrid inventions or applications of specific inventions, without imposing any a-priori constraint on the process of selection. In order to compare the IPC system and our classification, we will analyse the correspondence between our triples and IPC technological classes, by examining the distribution of each triple within IPC classes.

The empirical analysis is performed on patents applications at the European Patent Office between 1995 and 1999 (priority dates), whose abstract contains one or more

[8] The word technical implies that the invention must be usable in practice, in industry, and that it must also consist of more than the mere recognition of a law of nature (such recognition is called a (scientific) discovery and not a (technical) invention).

than one triple. Each patent is associated to a technological class, to a firm and to a country. In terms of technological classes, the relevant level of detail for our analysis is the *sub-class* at four or seven digit level (from now on we will call these sub-classes "technological classes"). Moreover firms are considered on the basis of their nationality, so that, for example, Ericsson US and Ericsson Sweden appear as two different firms. Different firms and establishments belonging to the same company within each country have been considered as a single unit. The ratio behind this choice lies in the possibility of analysing country-specific trends, which would be unfeasible if one considered the level of firm group.
In the data set a country may be assigned to each patent according to the nationality of the first inventor, of the patenting company or to the nationality of the industrial groups. Given the three possibilities of patent classification by country, we will concentrate on the eleven most innovative countries in terms of patenting activity (see Table 4 below). These countries account for almost 94% of the total patents. Our final sample consists of 20284 patents, 2708 firms and 70 technological classes at 4 digit level.

4.2 The characteristics of the triples

Among the triples extracted from patents' abstracts, we can identify some broad categories. There are generic or specific technologies (e.g. *digital signal processing*, *digital subscriber line, asynchronous transfer method*), technological platforms (e.g. *mobile communication systems*, *set top box*, *local area network*), products (e.g. *silicon nitride film, cathode ray tube*). The triple may define the technological dimension of the innovation or either the field of application of the patent considered.

Out of 119 triples, 4 appear in more than 1000 patents, and 8 in more than 500. Five technologies stand out as the most frequent ones (frequency in brackets): *mobile communication system* (1808), *data processing means* (1113), *printed circuits board* (1069), *composite video signal* (1001), *recorded recording medium* (998). It is possible to observe that the most frequent triples are related to the field of data transmission and to the area of audio-visual products/applications and include all the first three categories. Table A3 in the Appendix shows the distribution of the triples according to their dimension measured by the number of patents.

In order to compare the IPC classes and the triples, we analyse the degree of hybridisation of each triple, i.e. their distribution across different technological classes. In particular, we compute the number of technological classes in which patents containing a specific triple appear.

The triples expand over a wide range of IPC technological classes. We can identify 70 technological classes which contain at least one triple. Some of them (16) contain 30% or more of the total triples and are listed below. Out of these sixteen, six are in the H04 technological class (electric communication technique), two are in the G06 (computing; calculating; counting), three in the G01 (measuring; testing), two in the H01 (basic electric elements), two in the G11 (information storage) and one in the G09 (educating; cryptography; display; advertising; seals). It is worthwhile noting that this set of classes (see Table A4 and A5 in the Appendix) is much broader than the one typically used in standard statistics[9] (e.g. Eurostat, 2003). For each triple it is possible to calculate the concentration indexes of IPC classes. In particular we display the results for the C1 index which is the share of patents of the most important four-digit IPC class within each triple. Results for the C3 index at the four digit level or for the C1 at the seven digit level display the analogous patterns.

Using the C1 index it is possible to draw a comparison between IPC and the system of triples. The higher the C1, the higher is the degree of overlap of an IPC class over a selected triple and, accordingly, the more accurate is the contribution provided by the IPC class. Conversely low C1 show that the relevant triple is distributed across different technological classes and cannot be identified with one single class. Table 1 show the distribution of the C1 indexes over the different triples and shows that only in a limited amount of cases (21) the share of patents belonging to the most important IPC class is above 80%.

Table 1 Frequency of concentration ratio

Concentration ratio (C1)	
C1<0.25	20
0.25≤C1<0.50	48
0.50≤C1<0.80	31
C1≥0.80	21

Looking at the concentration ratio, it is also possible to detect more closely the characteristics of the selected triples. Triples with a high concentration rate very often identify very specific products: for example, *lithium secondary battery*, *gate insulating film*, *cathode ray tube* respectively record a C1 of 0.96, 0.99 and 0.79. On the contrary, triples with a low concentration rate generally correspond to technological platforms, general purpose technologies or very general products: for example, *digital signal processing*, *electronic control unit* and *high speed data* respectively display a C1 of 0.18, 0.19, 0.20. At an average level of concentration we find specific technologies or platforms: for example *asynchronous transfer mode* and *projection optical system* display a C1 of 0.54; *graphical user interface* has a C1 of 0.57.

Table 2 examines the top ten triples by number of patenting firms. Despite an obvious correlation between the number of patents related to a specific triple and the number of patenting firms, Table 2 shows also that the amount of patents per firms may vary considerably across triples and that patenting activities in the different triples may be characterised by different degree of concentration at the firm level.

Table 2 Patents and patenting firms in selected triples

Triple	Number of	Number of

[9] If we consider a higher level of the IPC (three digit), it is possible to observe that the most relevant IPC classes are H04, H01, G06. Also G01 records the second highest number of triples, although it does not display a very high number of patents as compared to others. Table A5 in the Appendix lists all the technological classes in which we can find a triple (i.e. a relevant ICT-based product, technology or platform). Again this is an important insight even for research based on the IPC classes, since it allows to enlarge the set of classes usually considered when performing analysis on ICT patents (G06, H03 and H04).

	patenting firms	patents (rank)
Printed circuits board	434	1069 (3)
Data processing means	388	1113 (2)
Composite video signal	216	1001 (4)
Mobile communication system	215	1808 (1)
Reflected light beam	185	417 (10)
Data stored memory	181	468 (9)
Electrically conductive material	181	280 (16)
Power supply voltage	170	389 (11)
Central processing unit	169	379 (12)
Radio communication apparatus	168	619 (8)

Table 3 presents the most innovative firms in terms of number of patents. Among the top twenty firms, eight are from Japan, five from the US, one from Korea, two from Finland, one from Sweden, one from Germany, one from France and one from the Netherlands. Column 3 shows the amount of triples over which firms activities are distributed and column 4 shows the most important triple in terms of amount of patents (ST: specialisation triple). Column 5 shows the amount of patent (and the percentage share over the total amount of firms' patents) in the specialisation triple. Most of these firms patent mostly in the areas of m*obile communication system* and *composite video signal*, although they appear to be quite diversified across triples.

Table 3 The most innovative firms

FIRM	*PATENTS*	*TRIPLES*	*SPECIALISATION TRIPLE (ST)*	*N° OF PATENTS IN ST (%)*
NEC (JP)	967	83	Mobile communication system	106 (10.96%)
Matsushita Electronics (JP)	781	72	Recorded recording medium	108 (13.83%)
Sony (JP)	747	65	Recorded recording medium	225 (30.12%)
Siemens (DE)	601	73	Mobile communication system	107 (17.80%)
Lucent Technologies (US)	559	73	Mobile communication system	159 (28.44%)
Ericsson (SE)	507	61	Mobile communication system	134 (26.43%)
Nokia Networks (FI)	409	40	Mobile communication system	205 (50.12%)
Samsung Electronics (KR)	379	53	Mobile communication system	83 (21.90%)
Philips (NL)	368	66	Composite video signal	30 (8.15%)

Canon (JP)	363	49	Composite video signal	150 (41.32%)
Toshiba (JP)	344	61	Composite video signal	34 (9.88%)
Motorola (US)	327	64	Mobile communication system	55 (16.82%)
Texas Instruments (US)	318	55	Data processing means	39 (12.26%)
International Business Machines (US)	311	55	Data processing means	78 (25.08%)
Fujitsu (JP)	299	51	Projection optical systems	30 (10.03%)
Alcatel (FR)	290	52	Mobile communication system	57 (19.66%)
Hitachi (JP)	287	60	Recorded recording medium	36 (12.54%)
Nokia Mobile Phones (FI)	273	42	Mobile communication system	96 (35.16%)
Sharp (JP)	271	51	Composite video signal	40 (14.76%)
Ericsson (US)	222	36	Mobile communication system	60 (27.03%)

Table 4 illustrates some general features of the selected technologies as compared to the total sample in terms of geographical distribution of patenting activities. We use three different ways of assigning a country to a patent: country of residence of the first inventor, country of the applicant company, country of the industrial group. In terms of geographical distribution at the firm level, it is worth noticing that in both groups Japan and US account for the majority of the patents, with a respective share of 27.17 and 31.76 in the total sample, 36.02 and 30.28 in the sample of selected triples. However, US show a lower share of patents in the triples, while Japan displays a substantially higher share of patents in the selected technologies. Together with Japan, Sweden, Finland and South Korea show a positive difference between the share of patents in the selected technologies and the share of patents in the total sample. The result reflects the good performance of firms in these four countries in the production of ICT-related products and applications, as we showed above. Sweden (Ericsson) and Finland (Nokia) are particularly active in the mobile communications area as well as South Korea (Samsung), while Japan is very strong across different technological areas and hosts the top three firms in terms of patenting activity (NEC, Matsushita, Sony). Germany and France, on the contrary, display a lower share of patents in the selected triples, notwithstanding the presence of Siemens and Alcatel among the top innovators.

Table 4 Geographical distribution of patenting activity by group, firm and inventor

Whole sample				***Selected Triples***			
Country	*Group*	*Firm*	*Inventor*	*Country*	*Group*	*Firm*	*Inventor*
US	26.55	31.76	30.74	US	26.14	30.28	28.94
JP	32.46	27.17	25.75	**JP**	**36.91**	**36.02**	**36.03**
DE	13.09	14.02	15.27	DE	9.28	8.85	9.72
FR	8.20	7.07	6.70	FR	5.92	4.40	4.17
GB	1.67	3.18	4.59	GB	2.02	2.53	4.31
IT	0.40	1.90	2.21	IT	0.65	1.48	1.58

SE	3.63	2.29	2.06	**SE**	**5.29**	**3.05**	**2.31**
NL	6.63	2.74	1.83	NL	5.27	2.45	1.72
FI	2.64	2.08	1.81	**FI**	**4.69**	**4.02**	**3.57**
CH	2.03	1.62	1.45	CH	1.23	0.80	0.70
KR	0.35	1.43	1.43	**KR**	**0.32**	**2.60**	**2.52**
Patents	65746	103479	102547		16213	20284	20284
	(c)	(a)			(c)	(a)	(b)

(a) In case of co-patenting each country has been credited the patent
(b) The same patent may be counted in different triples
(c) Only for selected industrial group

If we look at the characteristics of patenting activity across country by group, it is possible to observe that the share of the US drops significantly: in terms of selected triples, this is probably due to the intense patenting activity of US subsidiaries of firms whose headquarters are located elsewhere (as in the case of Ericsson US, which ranks within the top 20 innovative firms). Indeed Sweden records an increase in the share of patents by group, most of which can be reasonably attributed to the strength of Ericsson worldwide. For the same reason, the share of Finland is higher when considering the level of group. An interesting result concerns the Netherlands, which goes from a share of 2.45 at the firm level, to a share of 5.27 at a group level.

5. *The nature of the emerging technologies*

This section is devoted to enquire the nature of the emerging technologies and to the presentation of comparative empirical evidence on emerging vs. non emerging technologies. From this point on, our *triples* are defined as *technologies*. We define p_{it} as the amount of patents in technology i at time t (i=1,….,119 selected technologies; t=95,…..,99 years). Technology i is an emerging technology, if its growth rate in terms of patents between the period (95-96) and the period (98-99) is above the average of the sample which includes all the technologies. We considered the sum over two years two avoid peaks due to random factors affecting the patenting procedures. The dummy variable EM in the table in the Appendix is equal to 1 for the emerging technologies and equal to 0 for the non emerging ones. The growth rate of the whole sample is 24.87%. Out of 119 triples, 58 are emerging and 61 are non emerging: the average growth rate of the emerging technologies is 71.64%, while the average growth rate of the non-emerging technologies is 0.63%. Although the rate of growth is not precisely a proxy for the emergence of new technologies, but more an indication of the importance of new or existing technologies, it is important to remember that our initial sample consists of patents from very recent years and, in this respect, it is possible to relate the emerging nature of a technology with its potential to generate new applications, products and technologies in different fields.

Table 5 provides some summary results comparing emerging and non emerging technologies in relation to the following issues: concentration of patenting activity across countries; concentration across IPC technological classes; concentration of patenting activity across firms.

5.1 Concentration of patenting activity across countries

The theoretical part of this paper suggests that innovative products and applications based upon emerging technologies tend to be more concentrated in a limited number of countries, because of the creation of locational and first mover advantages, as well as because of economies of scale and scope in the activities of research and production. Only in a second phase, when technologies are more standardized, innovative applications may be expected to diffuse across a wider number of countries. Here we test whether, among the selected technologies, there is evidence of a higher degree of geographical concentration in the emerging ones.

Table 5 Average Herfindahl indexes for emerging and non emerging technologies

		Number	NH_{FIRMi}	H_{FIRMi}	NH_{CLASSi}	H_{CLASSi}	*$NH_{COUNTRYi}$*	*$H_{COUNTRYi}$*
TOTAL		119	0,044	0,049	0,180	0,184	*0,338*	*0,342*
							NON EMERGING	
							0,3345	*0,3381*
NON EMERGING	*ALL*	*61*	*0,0400*	*0,0455*	*0,1163*	*0,1212*	*EMERGING*	
	DE	27	0,1599	0,1851	0,0804	0,1080	*0,3428*	*0,3468*
	FI	27	0,4794	0,5500	0,2306	0,3522		
	FR	27	0,2201	0,2682	0,1608	0,2135		
	JP	27	0,0844	0,0923	0,1146	0,1223		
	SE	27	0,6371	0,7018	0,1322	0,2858		
	US	27	0,0522	0,0618	0,1044	0,1133		
EMERGING	*ALL*	*58*	*0,0473*	*0,0532*	*0,2493*	*0,2538*		
	DE	21	0,3084	0,3335	0,3385	0,3630		
	FI	21	0,4327	0,4435	0,3937	0,4083		
	FR	21	0,3893	0,4252	0,2714	0,3072		
	JP	21	0,1273	0,1338	0,2551	0,2605		
	SE	21	0,7325	0,7420	0,3578	0,3860		
	US	21	0,0904	0,1001	0,2012	0,2096		

Let M_i and n_i be the total number of countries and the total number of patents applied for in technology i. In particular p_{im} is the total number of patents related to technology i, belonging to country m ($p_i = \sum_m p_{im}$; $m=1,..,M_i$). Accordingly $s_{im} = p_{im}/p_i$ is the share of country m of technology i patents. We can then build two Herfindahl indexes:

$\mathbf{H_{COUNTRYi}=\sum_m s^2_{im}}$ This is the Herfindahl index which illustrates the geographical concentration of patents for each technology i .

$\mathbf{NH_{COUNTRYi}=(n_i H_{COUNTRYi}-1)/(n_i -1)}$. This is the normalised Herfindahl index, which corrects the upward bias in the Herfindahl indexes based on count data, when the sample size is small (Hall, 2000).

Finally we can calculate the weighted average of the Herfindahl indexes for emerging and non emerging technologies (using as weights the dimension of the technology in terms of patents). Results are displayed in Table 7. Emerging technologies display a higher concentration across countries. This supports the idea that new inventions and applications related to emerging technologies are relatively more concentrated across country.

5.3 Concentration across IPC technological classes

The theory suggests that the amount of applications of a specific technology expands over time across different technological fields. Accordingly, as stressed in section 2, we may expect applications and inventions in emerging technologies to be relatively more concentrated in specific technological classes. As a consequence, we may expect to find a relatively higher amount of “hybrid” activities in established technologies. This issue can be investigated by asking whether inventions and new products related to a specific technology are concentrated in a specific technological class or, conversely, the technology constitute a fundamental base for a wide range of activities and, accordingly, related inventions are distributed over a large number of technological classes. To perform this analysis, we use IPC classes at 7 digit level, which is a very detailed level of classification.

Our selected technologies identify a set of n_i patents falling into k_i IPC classes. In particular, p_{ij} is the total number of patents related to technology i, belonging to the IPC class j ($p_i = \sum_j p_{ij}$; j=1,..,k_i). Accordingly $v_{ij} = p_{ij}/p_i$ is the share of technology i patents belonging to class j. We can then build two indexes:

$\mathbf{H_{CLASSi}= \sum_j v^2_{ij}}$ This is the Herfindahl index which illustrates the concentration of patents across specific IPC classes for each technology i.

$\mathbf{NH_{CLASSi}=(n_i H_{CLASSi} -1)/(n_i -1)}$ This is the corrected Herfindahl index (Hall, 2000).

Finally we have calculated the simple average of the two Herfindahl indexes for emerging and non emerging technologies, always controlling for the size of the technology in the sample. Herfindahl indexes are calculated first for all the countries (ALL) and then for the top six countries in terms of innovative activity within the 119 technologies (Japan, US, Germany, France, Sweden, Finland). Results are displayed in Table 5.

First, average Herfindahl indexes display a statistically significant difference between emerging and non emerging technologies. Emerging technologies are more concentrated in some IPC classes as compared to non emerging. This supports the idea that technologies in their initial phase are embedded and used in applications and inventions that are relatively more concentrated in specific technological classes and that the technological domain of a specific invention is quite limited. The most significant result however concerns the invariance of this difference across countries: the concentration of emerging technologies in IPC classes is much higher than for the non emerging ones in all countries considered.

5.4 Concentration of patenting activity across firms

Our theoretical framework does not give strong predictions on the relationship between concentration of innovative activities and technological evolution. On the one hand, evidence may suggests that inventions and new applications in emerging technologies can occur within very large, R&D intensive, multi-technology firms through a process of diversification and continuous innovation. On the other hand, the product life cycle approach emphasises the role of many innovative firms trying to catch new opportunities before the industry shake-out occurs.

Our selected technologies are a set of n_i patents applied for by Z_i firms. In particular p_{iz} is the total number of patents related to technology i applied for by firm z ($p_i = \sum_z p_{iz}$; $j=1,..,Z_i$). Accordingly $w_{iz} = p_{iz}/p_i$ is the firm z's share of patents related to technology i. In order to investigate concentration at the firm-level, we can again build two Herfindahl indexes:

$\mathbf{H_{FIRMi}= \sum_z w^2_{iz}}$ This is the Herfindahl index which illustrates the concentration of patents across countries for each technology i.

$\mathbf{NH_{FIRMi}=(n_i\, H_{FIRMi} -1)/(n_i -1)}$ This is the normalised Herfindahl index (Hall, 2000).

Finally we have calculated the weighted average of the two Herfindahl indexes for emerging and non emerging technologies. Results are displayed in table 5.

Also in this case, average Herfindahl indexes display a statistically significant difference between emerging and non emerging technologies. Emerging technologies are more concentrated in some firms and this supports the idea that there is a core of firms promoting innovations in emerging technologies. It is quite likely (see also Table 3) that firms developing innovations related to emerging technologies are a few large firms which, as Patel and Pavitt (1997) underline, are typically multi-field and differentiated. Once again, the higher concentration of emerging technologies at the firm level is invariant across country, suggesting that in each country emergent technologies are associated with higher concentration at the firm level. At the same time note that the country variance in terms of concentration is mainly due to countries' size.

The results described above receive even more strength if we look at the correlation between the rate of growth of the technologies and the concentration at the level of technological class, firm and country (Table 6). In particular, we can observe a positive and statistically significant correlation between the concentration of patents at the level of technological class and the concentration of patents at the level of firm, as well as a positive and significant correlation between the rate of growth of a technology and the concentration across technological classes. This means that emerging and important technologies tend to be associated with the research activity of a few firms and related to specific technological domains. Whether these domains are affected by firms' principal products as argued in Pavitt and Patel (1997), or are instead the result of an R&D activity directed at exploring new areas, is an open question and would require further investigation. The correlation between the concentration at the firm level and the growth rate is negative but statistically not significant.

Table 6 Correlation matrix

	NH_{FIRMi}	NH_{CLASSi}	$NH_{COUNTRYi}$	*GROWTH*
NH_{FIRMi}	1			
NH_{CLASSi}	**0,32249612**	1		
$NH_{COUNTRYi}$	**0,517015214**	0,201931683	1	
GROWTH	**-0,002161048**	**0,241938781**	0,003388871	1

6. A preliminary analysis on the knowledge sources using patents' citations

This section provides a preliminary attempt to use patent citations in order to detect the sources of knowledge and to understand their characteristics with reference to the selected technologies. Patent citations can be used to measure knowledge flows because of their legal dimension. Patent citations limit the scope of the property right in the patent claims[10] and represent an important linkage between applicants, their technological fields and their locations. The citation is used by the applicants of a specific patent to refer to a piece of previously existing knowledge: this implies that the specific patent builds upon the cited ones (Jaffe et al., 1993). As a result we can consider citations as a proxy for knowledge sources. Here, the preliminary analysis is aimed at verifying some stylised facts about knowledge sources in emerging and non emerging technologies. This is done along three dimensions:

- The degree of dispersion of knowledge sources across technological classes
- The degree of dispersion of knowledge sources across firms and/or the degree of self citations
- The role of universities and public research centres

The dispersion of knowledge sources by technological class and firm is calculated by using the index of originality (Trajtenberg et al. 2002; Hall 2000).

If c_{ij} is the total number of cited patents from technology i, belonging to the IPC class j ($c_i = \sum_j c_{ij}$; j=1,..,k_i; k_i is the amount of cited IPC classes). Accordingly $v_{ij} = c_{ij}/c_i$ is the

[10] As emphasised by Trajtenberg (1990), the list of citations is generated through a process involving the applicant, his attorney and the examiner that generates "*the right incentives to have all the relevant patents cited, and only those*" (p.174). This is particularly true for the EPO dataset. A survey of inventors by Jaffe et al. (2000) shows that citations can be used to track knowledge flows. Citations are a noisy signal for spill-over, but they show that the likelihood of knowledge spillover is significantly higher, if there is a citation. At the same time other studies show that patent citations are related to the value of the innovations (in terms of variation of a social surplus function) and to financial market valuation of the firms who own the patent (Trajtenberg, 1990; Hall et al. 2000).

share of citations from technology i belonging to class j. The corrected index of originality is therefore:

$ORIG_{CLASSi}$= (c_i /(c_i -1))*(1- $\sum_j v^2_{ij}$) The originality index illustrates how broad are the technological roots of the performed research. Table 7 illustrates the results. Self citations at the firm level are excluded from the calculation.

Investigating whether the sources of knowledge for the patents in the selected technologies belong to the same technological class allows to detect whether these technologies draw knowledge from a specific technological domain, or rely upon a wider set of technological fields and knowledge bases. The index of originality in this case is computed at the seven digit level. The results show that the sources of knowledge in the emerging technologies seem to be relatively more concentrated in the same IPC classes than non emerging technologies (0.8112 vs. 0.8511), meaning that they relate to a less broad technological domain and rely upon a relatively more specific sources of technological knowledge.

Table 7 Dispersion of knowledge sources

	$ORIG_{CLASS}$	***$ORIG_{FIRM}$***	***NON FIRM PATCIT***	***SELFCIT***
Total sample	0.8317	0.9572	12.7141	16.7177
Non emerging	0.8511	0.9541	11.1659	18.9591
Emerging	0.8112	0.9604	14.3424	14.3603

Examining the dispersion of knowledge across firms provides interesting insights both on the nature of technological linkages between different firms patenting in the ICT area and on the existence of more or less intense flows of knowledge across different actors.

If c_{iz} is the total number of cited patents from technology i, applied for by firm z ($c_i = \sum_z c_{iz}$; z=1,..,Z_i). Accordingly $w_{iz} = c_{iz}/c_i$ is the share of citations from technology i to patents applied for by firm z. In this case the corrected index of originality is therefore:

$ORIG_{FIRMi}$= (c_i /(c_i -1))*(1- $\sum_z w^2_{iz}$) The originality index illustrates how broad are the knowledge sources in terms of amount and concentration of predecessors involved in previous research. Self citations at the firm level are excluded from the calculation.

The empirical findings reveal that the technological and knowledge base of emerging technologies is more dispersed across firms as compared to non emerging ones. This means that, although a few firms account for most of the innovative activity related to the emerging technologies in the ICT area, these firms draw knowledge and technological sources not only from internal, but also from external resources.

Moreover we have calculated for each technology i the share of self citations at the firm level divided by the total amount of citations of firms in technology i. We can observe that self-citations are more common in non emerging as compared to emerging technologies, which reinforces the idea of firms relying upon other actors in the developing of innovations. The ICT involves a number of extremely differentiated technological domains and the development of innovation and new applications often needs the combination of different firm specific knowledge and competencies. This is witnessed also by the large number of technology-based alliances in the sector and explains the variety of knowledge sources across firms for emerging technologies.

Finally, after having analysed the role of firms, it is interesting to compare this with the role of universities and public research centres. In particular, we can look at the share of cited patents belonging to these two types of institutions. Table 7 shows that this share is higher in emerging technologies than in non emerging ones, although this results is not very robust because it's sensitive to the inclusion of few key technologies. This seems to suggest a role for universities and public research centres as sources of knowledge and technological domains for emerging technologies in ICT, although further research is required in this field.

The matrix of correlation of all the relevant variables (table A2 in the appendix) confirms the relations described above, by emphasising that those relations hold independently of the size of each technology (in terms of number of patents). The growth rate is positively correlated with concentration across technological classes and negatively correlated with originality by technological class and with self citations: this means that emerging technologies are characterised by dispersion of knowledge sources by class and by exchange of knowledge across firms. Concentration by firm is positively correlated with concentration by technological class and both of them are negatively correlated with originality of knowledge sources and self citations: this implies an important interplay between specificity of technological domain and knowledge sources, and existence of a core of innovative firms in emerging technologies. Finally, there is a positive correlation between originality by firm and originality by class, while both are again negatively correlated with self citations.

7. Conclusions

This paper develops a database of ICT-related technologies based upon patents' abstracts, in order to detect important technologies in the ICT area, without any subjective bias. This is done through the selection of the most frequent sequential triples of words and is controlled with a co-occurrence analysis of these triples within patents' abstracts.

These triples identify technologies/applications/products in the ICT fields that cut across many different IPC classes. Our research indicates that set of IPC classes involved by ICT innovative activities should be broader than the ones considered.

Among this set of technologies/applications/products this paper identifies the emerging technologies by looking at the rate of growth of patents, and analyses whether there is a distinct pattern in the development and exploitation of new technologies. On the one hand, it enquires the nature of the knowledge sources in these fields and investigates whether the technologies are related to a single domain within a selected and homogeneous set of technological principles or, instead, rely upon different principles in different technological fields. On the other hand, it illustrates the characteristics of the actors involved in the innovation process and in particular it analyses industrial concentration of innovative activities in the promotion of new technologies. In doing so, the paper distinguishes between emerging and non emerging technologies.

As far as emerging technologies are concerned, results show that innovation is more concentrated across firms and technological classes (and countries) in emerging technologies relatively to non emerging ones and that this result is invariant across countries. This supports the idea that there is a core of firms promoting innovations in specific technological domains, and preliminary analyses seem to suggest that these are large firms. In particular the analysis suggests that big firms are also diversified over a wide range of technological activities.

Citations by patents in emerging technologies appear also to be more concentrated as compared to non emerging ones in terms of technological classes. This supports the idea that technologies in their emerging phase are embedded and used to develop applications and inventions that are relatively more specific and related to well-defined technological domains. Accordingly, it is more likely to find hybrid activities with reference to established or non emerging technologies.

We also find that emerging technologies are less concentrated in terms of patents citations across firms, suggesting a dispersion of knowledge sources across innovative actors. Finally, universities and public research centres seem to play a role as a source of knowledge in emerging technologies, although this result is not very strong.

References

Corrocher N. and Montobbio, F. (2003), "Analysing the emergence of new industrial activities: some methodological issues". TENIA Working Paper.

Coulter, N., Monarch, I., and Konda, S. (1998), "Software engineering as seen through its research literature: A study in co-word analysis", *Journal of the American Society for Information Science*, 49(13), pp.1206-1223.

Courtial, J.P. (1994), "A coword analysis of scientometrics", *Scientometrics* 31(3), pp.251-260.

Courtial, J.P., Callon, M. and Sigogneau, A. (1993), "The use of patent titles for identifying the topics of invention and forecasting trends", *Scientometrics*, 26(2), pp.231-242.

Dasgupta, P. and Stiglitz, J. (1980), "Industrial Structure and the Nature of Innovative Activity", *Economic Journal*, 90(358), pp 266-293.

Davies, A. (1996), "Innovation in Large Technical Systems: The Case of Telecommunications", *Industrial and Corporate Change*, 5(4), pp.1143-1180.

Ding, Y., Chowdhury, G.C., and Foo, S. (2000), "Bibliography of information retrieval research by using co-word analysis", *Information Processing and Management*, 37, pp.817-842.

Ehrnberg, E. and Jacobsson, S. (1997), "Technological Discontinuities and Incumbents' Performance: An Analytical Framework", in Edquist, C. (ed.) *Systems of Innovation - Technologies, Institutions and Organizations*. Pinter, London.

Eurostat (2003), "Increase of patent applications to the EPO in the ICT sector between 1991 and 2001", *Statistics in Focus, Science and Technology*.

Fai, F. and von Tunzelmann, N. (2001), “Industry-specific competencies and converging technological systems: evidence from patents”, *Structural Change and Economic Dynamics*, 12, pp.141-170.

Fujimoto, M., Miyazaki, K. and von Tunzelmann, N. (2000), “Technological fusion and telemedicine in Japanese companies”, *Technovation*, 20, pp.169-187.

Gort, M. and Klepper, S. (1982) Time Paths in the Diffusion of Product Innovations, *The Economic Journal*, 92, pp. 630-653.

Granstrand, O., Patel, P. and Pavitt, K. (1997), “Multitechnology corporations: whey they have ‘distributed’ rather than ‘distinctive core’ competencies”, *California Management Review*, 39, pp.8-25.

Grossman, G. M. and Elhanan, H.(1991) Quality Ladders and Product Cycles, *The Quarterly Journal of Economics,* 106*,* pp. 557-86.

Hall B.H. (2000); A Note on the Bias in the Herfindahl Based on Count Data. Mimeo. UC Berkeley and Nuffield College, Oxford. http://emlab.berkeley.edu/users/bhhall/hhibias.pdf

Hall B.H., Jaffe A.B., Trajtenberg M. (2000) Market Value and Patent Citations: A First Look, NBER Working Paper No. 7741.

Helpman, E. (ed.) (1998), *General Purpose Technologies and Economic Growth*. MIT Press, Cambridge.

Jaffe AB, Trajtenberg M, Fogarty MS (2000) The Meaning of Patent Citations: Report on the NBER/Case-Western reserve survey of Patentees. NBER Working paper series n 7361.

Jaffe AB. Trajtenberg M Henderson R. (1993) Geographic Localisation of Knowledge Spillovers as Evidenced by Patent Citations. Quarterly Journal of Economics. 108: 577-598

Klepper S. (1996) “Entry, exit, growth and innovation over the product life cycle”, *American Economic Review*.

Klepper, S. (1996) Entry, Exit, Growth and Innovation over the Product Life Cycle, *American Economic Review*, 86,3, pp. 562-583.

Klepper, S. (1997) Industry Life Cycles, *Industrial and Corporate Change*, 6, 1, pp.145-181.

Klepper, S. and Graddy, E. (1990); The Evolution of New Industries and the Determinants of Market Structure, *RAND Journal of Economics*, v. 21(1), pp. 27-44.

Kodama, F. (1992), “Technology Fusion and The New R&D”, *Harvard Business Review*, July-August, pp.70-78.

Koumpis, K. And Pavitt, K. (1999), “Corporate Activities in Speech Recognition and Natural Language: Another ‘New-Science’-Based Technology”, *International Journal of Innovation Management*, 3(3), pp.335-366.

Krugman, P. (1979) A Model of Innovation, Technology Transfer, and the World Distribution of Income, *Journal of Political Economy*, 87, 21, pp. 253-66.

Krugman, P. (1995) Technical Change in International Trade, in Stoneman, P. (ed.) *Handbook of the Economic of Innovation and Technological Change,* Blackwell, Oxford, chapter 9, pp. 342-365.

Mahdi, S. and Pavitt, K. (1997), "Key National Factors in the Emergence of Computational Chemistry Firms", International Journal of Innovation Management, 1(4), pp.355-386.

Malerba, F. (2002) "Sectoral systems of innovation and production", *Research Policy*

Malerba F., Montobbio F. (2003); Exploring Factors Affecting International Technological Specialization: the Role of Knowledge Flows and the Structure of Innovative Activity, *Journal of Evolutionary Economics*, v. 13, n. 4, pp. 411-434

Mansell, R. and Steinmueller, W.E. (2000), Mobilizing the Information Society. Oxford University Press, Oxford.

Metcalfe S. (1998), Evolutionary economics and creative destruction. Routledge, London

Miyazaki, K. (1994), "Search, Learning and Accumulation of Technological Competences: The Case of Optoeletronics", *Industrial and Corporate Change*, 3(3), pp.631-654.

Nelson RR Winter S (1982) *An evolutionary Theory of Economic Change*. The Belknap Press, Cambridge (Mass).

Nelson, R.R. (1994), "The Co-evolution of Technology, Industrial Structure and Supporting Institutions", *Industrial and Corporate Change*, 3(1), pp.47-63.

Nelson, R.R. (ed.) (1993) *National Innovations Systems: A Comparative Study*, Oxford University Press, Oxford.

Noyons, E.C.M. and van Raan, A.F.J. (1998), "Monitoring scientific developments from a dynamic perspective: self-organised structuring to map neural network research", *Journal of the American Society for Information Science*, 49(1), pp.68-81.

Pavitt, K. (1994), "Key Characteristics of Large Innovating Firms", in Dodgson, M. and Rothwell, R. (eds.), *The Handbook of Industrial Innovation*. Elgar Publishing, Cheltenham, UK.

Patel, P. and Pavitt, K. (1997), "The technological competencies of the world's largest firms: complex and path-dependent, but not much variety", *Research Policy*, 36, pp.141-156.

Pistorious, C.W.I. and Utterback, J.M. (1997), "Multi-mode interaction among technologies", *Research Policy*, 26, pp.67-84.

Rikken, P., Jiers, H.A.L., and Vos, R. (1995), "Mapping the dynamics of adverse drug reactions in subsequent time periods using INDSCAL", *Scientometrics*, 33(3), pp.367-380.

Rip, A. (1995), "Introduction of New Technology: Making Use of Recent Insights from Sociology and Economics of Technology", *Technology Analysis & Strategic Management*, 7(4), pp.417-431.

Rosenberg, N. (1976), *Perspectives on technology*. Cambridge University Press.

Trajtenberg M (1990) A Penny for Your Quotes: Patent Citations and the Value of Innovations. *Rand Journal of Economics*, 21(1): 172-187

Trajtenberg M., Henderson, R. Jaffe A. (2002); University versus Corporate Patents: a Window on the Basicness of Invention, in Jaffe A. Trajtenberg M.; Patents, Citations & Innovations. MIT Press, Cambridge, Ch. 3.

Tushman, M.L. and Anderson, P. (1986), "Technological Discontinuities and Organisational Environment", Administrative Science Quarterly, 31, pp.439-465.

Utterback, J. M. (1994), *Mastering the Dynamics of Innovation*, Harvard Business School Press, Boston.

Van den Ende, J. and Kemp, R. (1999), "Technological transformations in history: how the computer regime grew out of existing computing regimes", Research Policy, 28, pp. 833-851.

Van Raan, A.F.J. and Tijssen, R.J.W. (1993), "The neural net of neural network research", *Scientometrics*, 26(1), pp-169-192.

Vernon, R. (1966) International Investment and International Trade in the Product Cycle, *The Quarterly Journnal of Economics*, 80, 2 , pp. 190-207.

Vernon, R. (1979) The Product Life Cycle Hypothesis in a New International Environment, *Oxford Bulletin of Economics and Statistics*, 41, 4, pp 255-67

Zucker L.G., Darby M.R., Armstrong J.(1994), "Intellectual capital and the firm: the technology of geographically localized knowledge spillovers", NBER Working Paper No. 4946, December.

APPENDIX

A1 A Sample of selected technologies

TRIPLE	*EM*
Local Area Network	0
Asynchronous Transfer Mode	0
Communication Base Station	1
Code Division Multiple	1
Mobile Communication Systems	1
Central Processing Unit	0
Satellite Communication System	1
Composite Video Signal	0
Connected Input Terminal	0
Data Communication System	1
Data Processing Means	0
Recorded Recording Medium	1
Data Stored Memory	1
Method Transmitting Data	1
Digital Audio Signal	1
Digital Data Stream	0
Digital Signal Processing	0
Projection Optical System	1
Plasma Processing Chamber	1
Liquid Crystal Display	0

A2 Correlation matrix

	GROWTH	*SIZE*	NH_{FIRM}	NH_{CLASS}	$ORIG_{CLASS}$	$ORIG_{FIRM}$	*NON FIRM PATCIT*	*SELFCIT*
GROWTH	1							
SIZE	-0.0584	1						
NH_{FIRM}	-0.0022	-0.0884	1					
NH_{CLASS}	**0.2419**	-0.0844	**0.3225**	1				
$ORIG_{CLASS}$	**-0.1920**	0.0488	**-0.3382**	**-0.8752**	1			
$ORIG_{FIRM}$	0.0062	0.0516	**-0.6808**	**-0.4526**	**0.5424**	1		
NON FIRM PATCIT	-0.0033	-0.0559	-0.0644	0.0683	0.0318	0.0891	1	
SELFCIT	**-0.1116**	-0.0424	**0.5694**	0.2552	**-0.3783**	**-0.8620**	0.0111	1

A3 Frequency of patents within triples

Number of triples	*Frequency*
4	>1000
4	500 < x ≤ 1000
14	200 < x ≤ 500
31	100 < x ≤ 200
36	50 < x ≤ 100
30	30 ≤ x ≤ 50

A4 Most important technological classes (4 digit)

IPC Code	***Name***	***Percentage of triples***
H04B	TRANSMISSION	63,03%
H04N	PICTORIAL COMMUNICATION e.g. TELEVISION	63,03%
H04L	TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION	59,66%
G06F	ELECTRIC DIGITAL DATA PROCESSING	57,98%
H04Q	SELECTING	51,26%
G11B	INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER	47,90%
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR	46,22%
G01R	MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES	44,54%
G06K	RECOGNITION OF DATA; PRESENTATION OF DATA; RECORD CARRIERS; HANDLING RECORD CARRIERS	42,86%
G01N	INVESTIGATING OR ANALYSING MATERIALS BY DETERMINING THEIR CHEMICAL OR PHYSICAL PROPERTIES	39,50%
H01J	ELECTRIC DISCHARGE TUBES OR DISCHARGE LAMPS	35,29%
G01S	RADIO DIRECTION-FINDING; RADIO NAVIGATION; DETERMINING DISTANCE OR VELOCITY BY USE OF RADIO WAVES; LOCATING OR PRESENCE-DETECTING BY USE OF THE REFLECTION OR RERADIATION OF RADIO WAVES; ANALOGOUS ARRANGEMENTS USING OTHER WAVES	34,45%
H04J	MULTIPLEX COMMUNICATION	33,61%
G11C	STATIC STORES	31,93%
G09G	ARRANGEMENTS OR CIRCUITS FOR CONTROL OF INDICATING DEVICES USING STATIC MEANS TO PRESENT VARIABLE INFORMATION	31,09%
H04H	BROADCAST COMMUNICATION	30,25%

A5 Most important technological classes (3 digits)

IPC Class	***Number of triples***	***Number of patents***
H04	**93**	**8343**
H01	**90**	**4150**
G06	**79**	**2095**
G11	67	1784
G01	**91**	**1360**
H03	67	1324
H05	50	774
G09	47	454