Available online at www.sciencedirect.com
ScienceDirect
Research Policy 36 (2007) 418–432
research policy
www.elsevier.com/locate/respol

Schumpeterian patterns of innovative activity in the ICT field

Nicoletta Corrocher[a,*], Franco Malerba[a], Fabio Montobbio[a,b]

[a] *CESPRI, Bocconi University, Via Sarfatti 25, 20136 Milan, Italy*
[b] *Department of Economics, University of Insubria, Via Monte Generoso 71, 21100 Varese, Italy*

Received 14 March 2005; received in revised form 2 September 2006; accepted 10 January 2007
Available online 15 February 2007

Abstract

This paper studies the patterns of innovation in the ICT field using patents and patent citations. It provides an original methodology to identify ICT applications using patent abstracts and selecting the most frequent sequential triples of words without any a priori selection of keywords. This paper shows that the set of IPC classes related to ICT is broader than the one usually considered.

Moreover, our results show that ICT applications can be distinguished into two main groups in terms of growth and structure of innovative activities, technological pervasiveness, and knowledge sources. High opportunity ICT applications are characterised by high growth of patenting activity, high rate of entry of new innovators and high concentration of technological activity across firms. They also display a diversified knowledge base in terms of technological domains and actors involved. Conversely, low opportunity ICT applications are characterised by a lower growth and by a lower concentration of innovative activities across firms, as well as by a lower rate of entry of new innovators. Innovations in these ICT applications show less diversified knowledge sources and a higher degree of internal knowledge base.

Keywords: Innovation patterns; ICT; Patent abstracts search

1. Introduction

How can patterns of innovation in the ICT field be identified, given the large variety of applications, the diversity in the knowledge base among applications and the substantial heterogeneity of actors? This paper tries to address this question by providing an original methodology to identify ICT applications using patent abstracts and to examine the features of innovative activities employing a wide range of indicators.

ICT is a key sector in the economy and a major source of technical change and productivity growth. Its growing importance is reflected in the increasing number of patent applications as well as in its rising share in total patents (Eurostat, 2003).[1] However, for patent data, standard documents by Eurostat and OECD usually refer to

[1] A recent bulletin by Eurostat (2003) shows that, in 2001, the share of the 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% versus 11.0%), but also for the United States (22.0% versus 10.9%) and Japan (17.7% versus 11.9%).

* Corresponding author. Tel.: +39 02 5836 3396;
fax: +39 02 5836 3399.
E-mail addresses: nicoletta.corrocher@unibocconi.it (N. Corrocher), franco.malerba@uni-bocconi.it (F. Malerba), fabio.montobbio@uninsubria.it (F. Montobbio).

0048-7333/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2007.01.002

a selected number of International Patent Classification (IPC) classes,[2] which vary across different reports.

In this paper, we claim that the use of IPC classes to describe technological applications related to ICT – and to single out the patterns of innovative activity – presents major drawbacks, and we propose a novel methodology for identifying relevant ICT applications. Innovations in ICT may not be associated with existing technological classes because innovations in this field involves a very broad set of technological applications and technological progress in ICT proceeds in many different directions. In addition, the general purpose character of ICT is responsible for the emergence of ICT innovations in different technological domains, so that it is often difficult to assign a patent to an exogenously defined class. The methodology proposed identifies relevant ICT applications using patent abstracts. We show that ICT applications spread across a set of IPC classes, which is wider than the one commonly used.

In this paper, we also investigate patterns of innovative activities in the ICT field. We start from the basic distinction regarding Schumpeterian patterns of innovation (Malerba and Orsenigo, 1997), which is based on indicators such as concentration of innovative activity, stability in the ranking of innovators and entry of new innovators, and we enlarge it in several directions. First, in addition to firms, we consider other actors involved in innovation. Second, we look at the growth and technological pervasiveness of ICT-related innovations. Finally, we examine the variety of the sources of knowledge affecting patenting activities.

Our results show that ICT applications can be distinguished into two main groups in terms of growth of patents, structure of innovative activities, pervasiveness of innovations and variety of knowledge sources. High opportunity ICT applications are characterised by a high growth of patents, high rate of entry of new firms and high concentration of technological activity across firms. Innovations related to these applications are distributed over a few technologies, and have a high variety of knowledge sources in terms of technological domains and actors involved. Conversely, low opportunity ICT applications are characterised by a lower growth of patents, lower concentration of innovative activities across firms, and lower rate of entry of new innovators. Innovations in these ICT applications show a lower variety of knowledge sources and a higher degree of self-citations, but spread over a large number of technological classes.

The paper is structured as follows. In Section 2 we discuss some background literature on sectoral patterns of innovation in ICT at the qualitative level, by focussing on sectoral specificities, knowledge sources and actors involved in the innovation process. In Section 3 we discuss the data and the methodology. The methodology used to identify ICT-related applications involves the selection of relevant triples of words from patent abstracts, which allows detecting, without a subjective bias, all the relevant applications in the ICT area. Sections 4 and 5 contain the empirical analysis of the Schumpeterian patterns of innovation in ICT. Section 4 describes the variables used in the analysis and provides a preliminary descriptive analysis of innovation in ICT. Section 5 describes the two different types of patterns of innovation in ICT. Section 6 compares the methodology used in this paper with the same analysis conducted on standard IPC classes. Finally, Section 7 concludes and discusses some prospect for future research.

2. Sectoral patterns of innovation and ICT

The innovation process differs across sectors in terms of various dimensions, as Pavitt (1984), Mowery and Nelson (1999), Malerba (2002, 2004) among others have shown. In particular, the structure of innovative activity and the way in which technologies develop vary considerably across different industries and technological fields. The literature has underlined the existence of different patterns of innovation, providing widely used taxonomies (Kamien and Schwartz, 1982; Nelson and Winter, 1982; Pavitt, 1984; Malerba and Orsenigo, 1995, 1997). In particular, Malerba and Orsenigo (1997) propose the existence of different Schumpeterian patterns of innovation, which are basically the result of specific technological regimes in terms of opportunity and appropriability conditions and degree of cumulativeness. They single out two patterns. The first pattern (Schumpeter Mark I or "widening") is characterised by creative destruction: a high rate of entry of entrepreneurs and new firms in innovative activities, and a high turbulence in the hierarchy of innovators. The second pattern (Schumpeter Mark II or "deepening") is characterised by creative accumulation: a dominance of large established firms and a stability of the leading innovators. However, sectors differ also in the organisational rules and institutional arrangements according to which technical change takes place (Dosi, 1997;

[2] In particular, OECD (2002) includes patents from any of the following classes G06, G11, H04. Eurostat (2003) defines as ICT the following classes: G06, H03 and H04. See also the Patent Manual (OECD, 1994).

Malerba, 2002; Montobbio, 2003, 2004), as well as in terms of the characteristics of the knowledge base of innovative activities, which in turn implies different sources of information and different solving procedures. Finally, sectors vary in terms of the intensity and stability of firms' co-operation in R&D, types of interactions in (vertically or horizontally) linked sectors and, finally, firms' reliance upon formal scientific research performed in external research laboratories or universities.

Drawing from this literature, the paper aims to investigate the existence of distinct patterns of innovative activity in ICT applications. In doing so, the paper concentrates in particular on innovative actors (Section 2.1), degree of pervasiveness of innovations (Section 2.2) and variety of the knowledge sources for innovation (Section 2.3).

2.1. Actors involved in the innovative activity

The actors involved in innovative activities vary very much, from large incumbent companies to new firms, to other types of organisations such as universities and research laboratories. Often big established firms are major players in promoting new applications and inventions (Pavitt, 1994; Patel and Pavitt, 1997) and, up to a certain level, higher concentration stimulates innovation, because of the availability of big research teams, large internal financial resources and cumulativeness of learning and technical advance (Dasgupta and Stiglitz, 1980; Nelson and Winter, 1982; Malerba and Montobbio, 2003).

The technological life cycle model suggests the presence of other types of actors (Klepper, 1996). In the case of a radical innovation, the change in competences and organisation in established companies endowed with a specific and historically rooted set of skills may be a difficult and slow process, so that their ability to quickly respond to this type of change can be limited (Tushman and Anderson, 1986; Patel and Pavitt, 1997). Therefore, radical innovations may be associated with firms that are new and of small and medium size (Utterback, 1994). Large firms and industrial concentration appear in the market only when a dominant design has already emerged and the product is more standardised (Klepper, 1996).

Public research institutions and universities are also a major source of new technological and scientific knowledge (Nelson, 1993). The presence of excellence in science and universities appears to have guided the localisation of science-based industries: for example, the timing and location of new biotechnology firms is primarily explained by the presence of scientists who are actively contributing to basic science (Zucker et al., 1994).

Following the above discussion, we claim that within ICT the type of innovators may vary considerably across sectors and industry segments. Innovations may occur within very large, multi-technology firms, or within small firms or new innovators, or within universities and public research centres.

2.2. The technological pervasiveness of innovations

The recent literature has emphasised that technologies such as ICT spread across different industries (Kodama, 1992; Miyazaki, 1994; Koumpis and Pavitt, 1999; Fujimoto et al., 2000) because they are general purpose technology (Bresnahan and Trajtenberg, 1995; Helpman, 1998). General purpose technologies are characterised by pervasiveness of use and inherent potential for improvement and dynamism. In this respect, ICT has allowed the emergence of an increasing number of technologies and applications. Therefore, in this paper we are interested in investigating the extent to which ICT-related innovations spread across a wide range of sectors and technological domains.

2.3. The knowledge base and knowledge sources of innovations

As far as knowledge sources are concerned, the literature discussed above has suggested that new technologies may originate in various ways. They may stem from a single idea within a selected and homogeneous set of technological principles, or from the combination or integration of different pieces of knowledge. Empirical evidence shows that in ICT firms often develop applications using a highly diversified knowledge base (Koumpis and Pavitt, 1999; Mahdi and Pavitt, 1997), integrate new technologies and are technologically diversified (Granstrand et al., 1997). In addition, one of the main features of innovation in ICT is that it often stems from long-term R&D collaborations between firms, or between firms and other organisations across many different industries and research areas. Research capabilities and technological competencies outside a specific technology or sector may open important windows of opportunities for established companies or for new entrants.

In sum we claim that, in ICT high opportunity applications may rely upon a diversified knowledge base, but the degree of diversification of the knowledge base will greatly differ across applications.

3. Data and methodology

The analysis of patterns of innovative activity in the ICT field requires the identification of relevant technological applications. The current International Patent Classification (IPC) may not adequately capture all the relevant ICT applications. This is because the primary aim of the IPC is the establishment of an effective search tool. Therefore, the IPC classifies inventions according to their intrinsic nature (the "function-oriented" principle), rather than according to their possible applications. The function-oriented principle embraces 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.[3]

Because of the weaknesses of the IPC in detecting the nature and the relevance of ICT-related applications, in this paper we propose a different and more accurate methodology, which aims at identifying ICT applications without imposing any a priori constraint on the process of their selection. In particular, we examine the nature of ICT applications using the identification of relevant triples of words from patent abstracts.[4] Each patent document includes an abstract of the description of the invention, which outlines the existing technical background and knowledge (the *prior art*) on which the invention is based, as well as the contribution of the invention to the resolution of the technical problem involved. Patent abstracts are a reliable source of information, 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. We argue that different triples of words within patent abstracts can identify relevant technological domains (applications, platforms or products), which can be compared with the existing IPC technological classes. The count of patents containing the triple in their abstract can be used as a measure of innovative activity in that specific technological domain.

The recent literature on keywords analysis consists of extracting pre-selected words from keyword lists and titles. 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. However indexing might reflect the preconceptions and points of view developed by indexers during the course of their training and some inconsistencies in keyword selection by professional indexers working for different databases.

As a result we decided to follow a different route. We find words by directly extracting them from full-text documents. 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. This methodology allows the detection of inventions or applications of specific inventions, without imposing any a priori constraint on the process of selection.

3.1. Data

The empirical analysis is performed on patents and patent citations from the EP-CESPRI dataset. The EP-CESPRI dataset contains all the front page information about patent applications at the European Patent Office (EPO) since 1978. The EP-CESPRI database also contains 1,119,761 citations among EPO patents. Each patent is associated with a technological (primary IPC) class, to an applicant (firm or non-firm organisations) and to a country.[5] In terms of technological classes, we have a substantial level of detail, because we consider *sub-classes* at four- and at seven-digit levels (from now on we will call these sub-classes "technological classes"). Firms are identified 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.

[3] Some examples of function oriented classes are: transmission (H04B) or selecting (H04Q) (see also Table A2 in Appendix A). We call *applications* specific products and technologies like image processing system or mobile communication system.

[4] For related work on keyword and co-word analysis see Courtial et al. (1993), Van Raan and Tijssen (1993), Noyons and van Raan (1998), Ding et al. (2000). Courtial et al. (1993) in their co-word analysis use patents' titles that might be more affected by a specific strategic way of using words to distract patent examiners.

[5] A comparison between the distribution across countries of this sample and of the overall EPO dataset is analysed in Corrocher et al. (2003). We have shown that the distribution of ICT applications across countries are (patent shares in parenthesis): Japan (36%), US (30%), Germany (8.8%), France (4.2%), Finland (3.8%), Sweden (3%) and Korea (2.8%). However, in this paper, we do not consider here a particular selection of countries or data at country level. We know from previous work of some us (Malerba and Orsenigo, 1995, 1996, 1997; Montobbio, 2003) that sectoral patterns of innovation display strong similarities across countries. Moreover many actors in our sample operate on the technological applications in ICT at the international level.

We take into consideration the time period 1995–1999. Years are assigned to patents using the priority dates.[6]

We start from a wide set of technological classes that refer to ICT. In particular, we select eight broad technological classes at the three-digit level belonging to the sections of *Physics* (*G*) and *Electricity* (*H*) of the IPC (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.[2] Our final sample is made of 102,547 patent abstracts.

In order to detect the sources of knowledge and to understand their characteristics with respect to the selected ICT applications, we use patent citations. There is an expanding literature showing that patent citations can be used to measure knowledge flows.[7] Patent citations limit the scope of the property right and represent an important linkage among inventors/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).

3.2. Methodology

From this sample of patent abstracts, we identify relevant ICT applications without imposing any predetermined keyword.[8] This methodology includes three steps: the extraction of triples of words from patent abstracts; the selection of triples that appear with a significant frequency; the identification of relevant triples.[9] These triples constitute different types of technologies, applications, platforms or products in the ICT area. In order to simplify the terminology from this point on, we define our *triples* as ICT applications.

The first step involves the extraction of words from patent abstracts through an ad-hoc algorithm. We assume that a technological application can be adequately identified by a sequence of at least three words.[10] The algorithm is therefore developed with the aim of extracting triples of sequential words.[11] From the initial data set of 102,547 patent abstracts we have extracted 6,953,563 triples. The second step concerns the selection of triples occurring with a significant frequency. We have chosen triples that occur in at least 30 patents.[12] The third step identifies the triples, i.e. the relevant ICT applications.

[6] 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 for 2000 and 2001 are not used for this analysis.

[7] 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 European Patent dataset on which EP-CESPRI is based (Breschi and Lissoni, 2004). 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 spillovers 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).

[8] An alternative method would be to identify the occurrence of preselected ICT-related keywords in patents. However, as underlined by the literature and as emerged in some preliminary analysis (Corrocher and Montobbio, 2003), 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 patent abstracts, and even more, it does not allow to detect all the relevant technologies.

[9] Some caveats and limitations of the analysis need to 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, as 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 or specific use of words). We cannot control for all these aspects. However we can claim that the use of patent abstracts provides an improvement with respect to simple keywords (where the impact of strategic behaviour of applicant may be even stronger) and that patent abstracts are revised during the process leading to publication and therefore some language specificities are corrected.

[10] We choose to investigate triples of words, as previous analysis (Corrocher and Montobbio, 2003) aiming at the extraction of couples of words did not provide significant results.

[11] A preliminary extraction of triples has highlighted the existence of some problems in the procedure. First, there are many meaningless triples made of articles, prepositions, adverbs, verbs and so on: in order to solve this problem, an a priori "cleaning" of patent abstracts is required. 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"). We have taken into account these problems and proceeded with the cleaning of the abstracts and with a second extraction of triples of words.

[12] 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.

A relevant methodological issue concerns the handling of triples that represent the same application, but appear as different ones (e.g. radio base station and base radio station). In order to group triples that refer to the same application, we first analyse the complete dataset of triples to detect the most evident similarities, and then we perform a co-word analysis of triples within patent abstracts.[13] In particular we calculate the number of co-occurrences of triples within patent abstracts. The existence of co-occurrences between triples of words may identify a specific link that corresponds to a specific technology. Differently from the traditional literature, here the co-word analysis 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 technological application.[14] The technological meaning of the selected triples has been controlled with the help of engineers and experts in the ICT field.

The final result of this analysis is a list of 119 triples. The five most frequent triples are (number of patents in brackets): mobile communication system (1751), data processing means (1078), printed circuits board (1035), composite video signal (972), information recording medium (960) (for a broader selection, see Table A1 in Appendix A).[15] Our final sample consists of 19,353 patents, 19,187 citations to other patents, 2708 firms and non-firm organisations, 70 technological classes at four-digit level and 463 technological classes at seven-digit level.

[13] In doing this, we partially draw 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. Co-word analysis reveals patterns in a specific patent document (abstract) by measuring the strength of associations between terms. 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).

[14] For example, a frequent 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).

[15] Table A1 in Appendix A examines the top 10 triples by number of patents and patenting firms. Despite an obvious correlation between the number of patents related to a specific triple and the number of patenting firms, there is not a strict correspondence between the top ten triples in terms of patents and the top ten triples in terms of innovative firms.

3.3. IPC versus triples

The triples cover a wide range of IPC technological classes. We can identify patents in 70 technological classes at four-digit level. Some of these classes (16) contain 30% or more of the total triples and are listed in Table A2. Out of these 16, 6 are in the H04 technological class (electric communication technique), 2 are in the G06 (computing; calculating; counting), 3 in the G01 (measuring; testing), 2 in the H01 (basic electric elements), 2 in the G11 (information storage) and 1 in the G09 (educating; cryptography; display; advertising; seals). This set of classes is much broader than the one typically used in standard statistics (e.g. Eurostat, 2003).[16]

There is no one-to-one correspondence between the selected triples and the IPC classes. On the one side for each four-digit IPC class we can find many ICT applications. On the other side patents belonging to each triple may be classified in different IPC classes. In order to compare the IPC classes and the triples, we analyse the distribution of different technological classes within each triple. ICT applications expand over several four-digit IPC classes. On average, we can identify 14.6 classes per triple. Some triples however are much more dispersed across IPC classes. For example, patents related to data processing system, power supply unit, printed circuit board, response control signal, digital signal processing, light emitting element spread across more than 35 four-digit IPC classes.

In order to draw a comparison between IPC classes and the triples, we calculate the C1 index, i.e. the share of patents of the most important four-digit IPC class associated with each triple. The higher the C1, the higher is the degree of overlap between an IPC class and a selected triple and, consequently, the more accurate is the description provided by the IPC class. Conversely, a low C1 implies that the relevant triple is distributed across different technological classes and cannot be identified with one single class. The average value of C1 across triples is 0.51. This means that, on average, the most important IPC class accounts for just half of the patents in a specific triple. This indicates that in general the IPC is not accu-

[16] If we consider a higher level of IPC (three digit), it is possible to observe that the most relevant IPC classes are H04, H01, G06 (see Table A3 in Appendix A). Also G01 records the second highest number of triples, although it does not display a very high number of patents as compared to others. 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 (see footnote 2).

rate enough to capture the technological applications in the ICT field. Although some applications display a very high (>0.90) concentration rate – i.e. lithium secondary battery, fuel cell system, gate insulating film – and could be to some extent identified by using a single four-digit IPC, most triples record a very low concentration rate. As an example, digital signal processing, electronic control unit and high speed data display, respectively, a C1 of 0.18, 0.19, 0.20, so that they cannot be associated with a single technological class.

4. Empirical analysis: definition of the variables

Based on the discussion in Section 2, we have selected a set of variables aimed at identifying Schumpeterian patterns in ICT applications. We consider four dimensions: growth of patents, structure of innovative activities, technological pervasiveness, and knowledge sources.

4.1. Growth of patents

We investigate the growth of patents associated with different triples, in order to identify the rate of innovation in different ICT applications. We define P_{it} as the number of patents in triple i at time t and calculate the growth rate (GROWTH$_i$) of P_{it} between 1995–1996 and 1998–1999. We consider the sum over 2 years to avoid peaks due to random factors affecting the patenting procedures.

4.2. Structure of innovative activity

The structure of innovative activity is analysed in terms of concentration across firms and rate of entry.

4.2.1. Concentration of innovative activities across patenting firms

Let Z_i be the total number of firms, P_i the total number of patents applied for in triple i and P_{iz} the total number of patents in triple i applied for by firm z ($P_i = \sum_z P_{iz}$; $z = 1, \ldots, Z_i$). Accordingly $w_{iz} = P_{iz}/P_i$ is the firm z's share of patents in triple i. In order to investigate concentration at the firm-level, we build the Herfindahl index: $H_{\mathrm{FIRM}i} = \sum_z w_{iz}^2$.[17] Here we use the corrected version (Hall, 2000) that controls for the small sample bias:

$$\mathrm{CONC}_{\mathrm{FIRM}i} = \frac{(P_i\, H_{\mathrm{FIRM}i} - 1)}{(P_i - 1)}$$

[17] The Herfindahl index H refers to concentration and will be extensively used in the following analysis. In this case, it refers to the concentration of patents across firms for each triple i.

4.2.2. Entry of new patenting firms

In order to consider entry of new patenting firms, let Z_i be the total number of patenting firms and NZ_i the number of firms patenting for the first time in triple i. Then we define:

$$\mathrm{ENTRY}_{\mathrm{FIRM}i} = \frac{NZ_i}{Z_i}$$

This index describes the rate of technological net entry of firms in the patenting activity in triple i. Malerba and Orsenigo (1997) distinguish between net entry and lateral entry. The first refers to firms that innovate for the first time; the second refers to existing innovators that have already innovated in another technological class and diversify their activity.

Technological net entry can also be calculated in terms of the share of patents applied for by new innovators in triple i. If P_i and NP_i are the total number of patents and the number of patents applied for by new innovators in triple i, the following illustrates the share of new entrants associated with triple i:

$$\mathrm{ENTRY}_{\mathrm{PAT}i} = \frac{NP_i}{P_i}$$

4.3. Technological pervasiveness

The distribution of each technological application across IPC technological classes (at seven-digit and four-digit level) is analysed in order to capture the technological pervasiveness associated with each triple. Let K_i and P_i be the total number of IPC classes and the total number of patents applied for in triple i. In particular, P_{ik} is the total number of patents in triple i, belonging to the IPC class k ($P_i = \sum_j P_{ik}$; $k = 1, \ldots, K_i$). Accordingly, $v_{ik} = P_{ik}/P_i$ is the share of triple i patents falling into class k. We can then build the following indexes.

$H_{\mathrm{CLASS}i} = \sum_j v_{ik}^2$ is the Herfindahl index which illustrates the concentration of patents across specific IPC classes for each triple i. A low level of concentration indicates a high degree of technological pervasiveness of each triple i.

$\mathrm{HERF}_{\mathrm{TECH}i} = (P_i H_{\mathrm{CLASS}i} - 1)/(P_i - 1)$ is the corrected Herfindahl index (Hall, 2000). We distinguish between $\mathrm{HERF}_{\mathrm{TECH7}i}$ and $\mathrm{HERF}_{\mathrm{TECH4}i}$ depending on the level of aggregation of the IPC classes considered (number of digits).

4.4. Knowledge sources and related actors

The analysis of the knowledge sources in the selected triples is done along four dimensions: the variety of knowledge sources across technological classes, the variety of knowledge sources across firms, the importance of self-citations (internal to the organisation) and the role of universities and public research centres.

4.4.1. Variety of knowledge sources across technological classes

The relative variety of knowledge sources across technological classes is calculated using Herfindahl indexes (see also Trajtenberg et al., 2002 and Hall, 2000 for originality indexes). Let c_{ij} be the total number of cited patents from triple i belonging to the IPC class j ($c_i = \sum_j c_{ij}$; $j = 1, \ldots, K_i$; K_i is the number of cited IPC classes, at the seven-digit level). Accordingly $a_{ij} = c_{ij}/c_i$ is the share of backward citations from triple i belonging to class j.

We define the corrected Herfindahl index as: $\mathrm{HERFCIT}_{\mathrm{TECH}i} = (c_i \sum_j a_{ij}^2 - 1)/(c_i - 1)$. For each triple i, $\mathrm{HERF}_{\mathrm{TECH}i}$ and $\mathrm{HERFCIT}_{\mathrm{TECH}i}$ are highly correlated. The more a triple is dispersed across technological classes, the wider is the range of IPC classes over which knowledge sources (in terms of patent citations) expand. Since we want to analyse the variety of knowledge sources taking as given the characteristic of the triple in term of IPC composition, we have to define an index of relative concentration:

$$\mathrm{HERFSOURCES}_{\mathrm{TECH}i} = \frac{\mathrm{HERFCIT}_{\mathrm{TECH}i}}{\mathrm{HERF}_{\mathrm{TECH}i}}$$

This illustrates how varied the technological sources of the innovative activity are, conditional to the technological composition of the triple. Investigating whether cited patents in the selected ICT applications belong to the same technological class allows detecting whether these applications draw knowledge from a specific technological domain, or rely upon a wider set of technological fields and knowledge bases.

4.4.2. Variety of knowledge sources across firms

Examining the variety of knowledge sources across *firms* provides insights both on the nature of technological linkages between different firms patenting in the ICT area and on the existence of flows of knowledge across different actors.

If c_{iz} is the total number of cited patents from triple i, applied for by firm z ($c_i = \sum_z c_{iz}$; $z = 1, \ldots, Z_i$), $b_{iz} = c_{iz}/c_i$ is the share of backward citations from triple i to patents applied for by firm z. Self-citations at the firm level are excluded from the calculation. The corrected Herfindahl index is: $\mathrm{HERFCIT}_{\mathrm{FIRM}i} = (c_i \sum_z b_{iz}^2 - 1)/(c_i - 1)$. Again for the purpose of our analysis, we can define the following variable in relative terms:

$$\mathrm{HERFSOURCES}_{\mathrm{FIRM}i} = \frac{\mathrm{HERFCIT}_{\mathrm{FIRM}i}}{\mathrm{CONC}_{\mathrm{FIRM}i}}$$

This shows how varied the knowledge sources are in terms of firms involved in previous innovative activities.

4.4.3. Internal knowledge sources

We also calculate for each triple i the share of self-citations at the level of the individual organisation, i.e. the number of self-citations (sc_i) divided by the total number of backward citations (c_i). A self-citation occurs when the cited and the citing patents have the same applicant:

$$\mathrm{SELFSOURCES}_i = \frac{\mathrm{sc}_i}{c_i}$$

4.4.4. Universities and public research centres

Finally, we examine the role of universities and public research centres. In particular, we look at the share of cited patents belonging to these two types of institutions in each triple i, i.e. the number of citations to universities and public research centres (upc_i) divided by the total number of citations (c_i):

$$\mathrm{UPCSOURCES}_i = \frac{\mathrm{upc}_i}{c_i}$$

4.5. Descriptive analysis

Summary statistics for all these variables are displayed in Table 1.

The average number of patents per triple is 162.6. Out of 119 triples, 3 appear in more than 1000 patents and 8 in more than 500 (Table 2).

The average growth rate of patents in the triples is 43.4%, although there is a substantial variance across triples. For example, transmission power control and digital subscriber line display a growth rate (GROWTH) well above 200%, while private branch exchange, asynchronous transfer mode and signal input terminal show negative growth rates (GROWTH)—lower than −35%.

In terms of actors involved, Table A4 illustrates the most innovative firms in terms of number of patents. Column 2 displays the total number of patents by firm. Column 3 shows the number of triples over which firms activities are distributed. Column 4 shows the most important triple in terms of number of patents and its share over the total patents of each firm. Large innovators patent mostly in the areas of mobile communication system and composite video signal, although they have a diversified innovative activity across triples.

Table 1
Descriptive statistics for triples

	PATENTS	FIRMS	$CLASS_4$	TOTCIT	$C1_{FIRM}$	$C3_{FIRM}$	$ENTRY_{FIRM}$ (%)	$ENTRY_{PAT}$ (%)	GROWTH (%)
N	19353.00	2708.00	70.00	19187.00					
Mean	162.63	59.64	14.63	161.24	0.16	0.32	15.37	9.99	43.44
Median	82.00	35.00	13.00	69.00	0.12	0.28	14.85	8.91	20.42
Mode	32.00	30.00	5.00	20.00	0.10	0.19	0.00	0.00	0.00
S.D.	249.26	62.21	10.24	285.00	0.14	0.16	8.64	6.68	89.32
Variance	62130.64	3869.93	104.78	81222.22	0.02	0.03	74.73	44.67	7977.18
Minimum	28.00	1.00	2.00	15.00	0.04	0.09	0.00	0.00	−63.64
Maximum	1744.00	422.00	49.00	2183.00	1.00	1.00	40.00	31.25	657.14

	$HERF_{TECH7}$	$CONC_{FIRM}$	UPCSOURCES	SELFSOURCES	$HERFSOURCES_{TECH}$	$HERFSOURCES_{FIRM}$
N						
Mean	0.22	0.07	27.12	0.26	0.92	0.19
Median	0.13	0.04	24.66	0.00	0.85	0.14
Mode	0.02	0.01	0.00	0.00	0.26	0.00
S.D.	0.21	0.13	26.29	0.76	0.40	0.18
Variance	0.04	0.02	691.22	0.58	0.16	0.03
Minimum	0.02	0.01	0.00	0.00	0.26	0.00
Maximum	0.94	1.00	157.89	5.45	2.52	1.05

However the selected ICT applications do not appear to be particularly concentrated in terms of innovative activities. The average number of patenting firms in a triple is 60 and the concentration of innovative activity in terms of the share of the most innovative and of the three major innovators is low ($C1_{FIRM} = 0.16$; $C3_{FIRM} = 0.32$). More generally, concentration of innovative activities across firms measured by the Herfindahl index ($CONC_{FIRMi}$) is low. In particular, printed circuit board, data processing system, imaging processing system, mobile communication system have more than 200 patenting firms. Furthermore, there are 18 triples with patents from more than 100 firms and 48 triples with patents from more than 50 firms.

On average 15.4% of firms in our sample are totally new innovators, i.e. they started patenting in 1995 or afterwards in one of the selected triples. In terms of patents, 10% of all patents are applied for by totally new innovators. As underlined before, there is also an important role for new and possibly small firms in the ICT area. The percentage of new firms is particularly high (≥35%) in the following triples: worldwide web, graphical user interface, packet switched network.

Table 2
Frequency of patents (x) within triples

Frequency	Number of triples	Percentage
$x \leq 50$	35	29.4
$50 < x \leq 100$	33	57.1
$100 < x \leq 150$	19	73.1
$150 < x \leq 200$	11	82.4
$200 < x \leq 250$	3	84.9
$250 < x \leq 500$	10	93.3
$500 < x \leq 1000$	5	97.5
$1000 < x \leq 2000$	3	100.0

As far as the sources of knowledge are concerned, 96% of total cited patents belong to firms, while 4% belong to public research centres and universities. The average share of self-citations over total citations is relatively low (0.26%). Just nine triples display a percentage of self-citations higher than 1%. This means that, in general, sources of knowledge vary substantially (see Table 1).

5. Patterns of innovative activities

This section identifies the main characteristics of selected ICT applications in the ICT field. In order to identify patterns of innovative activities in the ICT field starting from the selected triples, we perform a cluster analysis. The first attempt results in three clusters. One of them is a small group composed by just three applications. A closer statistical investigation suggests that these three applications are characterised by a very high concentration of patenting firms and by a very high share of self-citations as compared to the rest of the sample. These applications are considered as outliers and are therefore eliminated from the sample.

The cluster analysis is then repeated on 116 observations. Two clusters of ICT applications emerge out of the analysis (Table 3) and can be associated with two different Schumpeterian patterns of innovations. In order to

Table 3
Clusters of triples

Variables	CLUSTER	
	High opportunity	Low opportunity
$ENTRY_{FIRM}$	1.0936	−0.37551
$ENTRY_{PAT}$	1.15731	−0.36874
GROWTH	0.68038	−0.22524
PATENTS	0.02549	0.00775
$HERF_{TECH}$	0.80619	−0.32639
$CONC_{FIRM}$	−0.03106	−0.18719
UPCSOURCES	0.30912	−0.11235
SELFSOURCES	−0.2496	0.01551
$HERFSOURCES_{TECH}$	−0.53308	0.18422
$HERFSOURCES_{FIRM}$	−0.14416	0.07892

investigate more precisely the characteristics of the two groups of applications, we test the equality of means, through an Independent Sample *T*-test (see Table A5 in Appendix A).

From the previous analysis it is possible to claim that innovative activities in the first Schumpeterian pattern are characterised by:

- high growth rate of patents,
- high concentration of patents across firms,
- high entry of totally new innovators,
- high number of patents applied for by the totally new innovators,
- low technological pervasiveness,
- high variety of knowledge sources across technological classes,
- high share of citations to patents from universities and public research centres,[18]
- low share of self-citations.[18]

Examples of these applications are: mobile communications systems, data communication system, packet switched network, optical transmission system, satellite transmission system.

Innovative activities in the second Schumpeterian pattern on the contrary display:

- low growth rate of patents,
- low concentration of patents across firms,
- lower entry of totally new innovators,
- limited number of patents applied for by the totally new innovators,
- high technological pervasiveness,
- low variety of knowledge sources across technological classes,
- low share of citations to patents from universities and public research centres,[18]
- high share of self-citations.[18]

Examples of these applications are: local area network, asynchronous transfer mode, information recording medium, liquid crystal display, random access memory.

We label the two distinct Schumpeterian patterns of innovative activity in the ICT field "high opportunity ICT applications" and "low opportunity ICT applications". For the scope of the present paper, opportunity conditions refer to the ease of innovation by firms and to the potential for innovation of each technological application (Malerba and Orsenigo, 1996), and are captured by the growth rate of innovative activities as measured by patents. Regarding the growth of patents, the first pattern displays a high growth rate, while the second one shows a lower growth rate.

In terms of structure of innovative activities, the "high opportunity" pattern is similar to a Schumpeter Mark I pattern, which is characterised by a major role of new innovators, but with an important addition. This pattern in ICT shows the significant coexistence of high entry of totally new innovators and high concentration. This means that high opportunity applications are associated both with the entry of new actors and with a core of firms promoting innovations. Compared to an archetypical Schumpeter Mark I pattern, this "high opportunity" pattern represents a more articulated structure, in which totally new innovators play a major role in giving dynamism to technological change in various applications, but they do so together with a core of incumbents.

The second pattern of innovative activities is more similar to a Schumpeter Mark II pattern, with a major role of existing firms in innovative activities, but also (and surprisingly) with low concentration of innovative activity across firms. This is not the archetypical Schumpeter Mark II pattern, because, in addition to a low rate of technological net entry, innovative activities are distributed across a large number of existing firms (rather than being concentrated in a few).

This analysis has also shed light on the technological *pervasiveness* of these two groups of ICT applications. Patenting activity in high opportunity applications takes place in specific technological classes. This is

[18] We check the robustness of the results through a two-sample Kolmogorov–Smirnov test, which does not assume normal distribution of variables. The test confirms the results, with the exception of the share of citations of public research centres and universities, and the share of self-citations (see Table A6 in Appendix A). In this case, the test cannot reject the null hypotheses of equality of means.

because high opportunity ICT applications are relatively more focussed on specific technological classes. On the contrary, innovations related to low opportunity ICT applications are distributed over a large number of technological classes.

The type and extent of *knowledge sources* represent another major distinction between high opportunity and low opportunity ICT applications. The sources of knowledge in high opportunity applications are more varied and spread across several IPC classes as compared to low opportunity applications. This means that innovative activities related to high opportunity applications stem from broad and heterogeneous technological domains. On the contrary, in the case of low opportunities applications, the sources of knowledge of innovative activities are less varied.[19] In addition, for lower opportunity ICT applications *self-citations* are relevant, implying also a more cumulative pattern of innovation carried out by the same set of established actors. High opportunity ICT applications, on the contrary, show a lower level of self-citations and a high share of citations of patents from universities and public research centres.[20]

In sum, a more articulated pattern of innovative activity compared to the archetypical ones discussed in Malerba and Orsenigo (1995, 1996) have emerged from the analysis of the ICT field, also because in this paper the technological pervasiveness and the variety of knowledge sources have been taken into account. High opportunity applications are more of a Schumpeter Mark I type, but here new innovators coexist with a significant number of large incumbents. Innovative activities take place in a few technological classes, but draw upon a wide variety of knowledge sources.

As we move from high opportunity to lower opportunity applications, the pattern becomes more of a Schumpeter Mark II, with a limited entry of totally new innovators. Again, in this case, a significant modification compared to the archetypical Schumpeter Mark II exists: the limited entry of totally new innovators is associated with a low concentration of innovators. In turn knowledge sources are less varied, but innovations take place in several technological classes.

[19] It must be noted that, when examining the dispersion of knowledge sources *across firms*, our analysis does not reveal, for high opportunity applications, any significant pattern in relation to the concentration or dispersion of the sources of knowledge.

[20] As stated in footnote 18, these last results are not very robust, since the non-parametric test rejects the hypothesis of significant differences between the two clusters in terms of self-citations and citations from universities and public research centres.

6. Does this methodology matter? A cluster analysis on IPC classes

The analysis conducted so far has expanded our understanding of patterns of innovation in ICT by identifying a set of technological applications which are different from the standard IPC classes because (as mentioned in Section 3) these classes do not correspond to the variety of applications in ICT.

In order to compare our methodology with the more traditional patent analysis based upon the IPC technological classes, we have examined the innovative activity in the ICTs field by performing a cluster analysis on the technological classes (at seven-digit level) that are usually taken as representative of the ICT area and belong to these broad categories: computing, calculating, counting (G06); basic electric circuitry (H03); electric communication technique (H04). The variables used for the cluster analysis are the same as before, although we could not calculate, for obvious reasons, the concentration of patents and citations across technological classes.

The cluster analysis (not reported here) does not reveal the existence of differences between IPC technological classes in terms of patterns of innovative activity in ICT. This is because IPC classes account for the intrinsic nature of the inventions more than for their possible applications. Since relevant ICT applications often stem from the combination of different bodies of knowledge and different technological backgrounds, the IPC proves to be not informative enough with respect to the characteristics of the innovative activity in the field. Therefore, we may confidently claim that the methodology presented here, by providing a finer grained analysis of applications in the ICT area, allows a much better understanding of innovative activities, Schumpeterian patterns and sources of knowledge in the ICT field than a similar use of IPC classifications.

7. Conclusions

This paper analyses patterns of innovative activity in ICT, by examining patenting activity and patent citations at the firm level for a large number of countries. In doing that, it introduces two novel features.

First, given that ICT has a wide variety of technological applications, the paper does not use the traditional IPC classification, but follows another route. It uses patent abstracts to detect important applications in the ICT field, without any subjective bias, by selecting the most frequent sequential triples of words, which identify technological applications in ICT. These applications cut across many different IPC classes. Our research therefore indicates that the number of IPC classes related to ICT applications is broader than the one usually considered.

Second, the paper enlarges the analysis of the Schumpeterian patterns of innovation, from the use of measures of concentration of innovative activities, stability of innovators and entry of new innovators, to other indicators such as the degree of pervasiveness of innovations across technological classes, the relevance of other actors (such as universities and research organisations) as sources of innovation, and the variety of knowledge sources at the technological level. In this way one could include in the analysis of Schumpeterian patterns of innovation non-firm actors whose role has become increasingly important, and some characteristics of the knowledge base.

The empirical evidence on the patterns of innovative activity in the ICT field shows that ICT applications can be distinguished into two main groups in terms of growth of patents and structure of innovative activities, technological pervasiveness, and variety of knowledge sources. High opportunity applications are characterised by a high growth of patents, high rate of technological net entry and high concentration of technological activity across firms. They show a varied and diversified knowledge base in technological terms and in terms of actors involved, but innovations are focused on a few technological classes. Conversely, low opportunity applications are characterised by a lower growth of patents and by a lower concentration of innovative activities across firms, as well as by a lower rate of technological net entry. Innovations in these ICT applications show less variety of knowledge sources and a higher degree of internal generation of knowledge, but they are distributed over a larger number of technological classes.

These results enrich our understanding of the Schumpeterian patterns by enlarging the number of variables under considerations and by applying the analysis to such an important area as ICT. For high opportunity applications, we indeed show a more articulated Schumpeter Mark I pattern, in which new innovators do not necessarily generate high turbulence in the industry, high innovative entry coexists with a certain level of concentration, and in which a variety of sources of knowledge coexists with innovations focused on a few technologies. Similarly, also for low opportunity applications a more articulated Schumpeter Mark II pattern emerges from our analysis, in which limited innovative entry is associated with a distributed innovative activity across firms, and in which low variety in the sources of knowledge is coupled with a large number of technological classes in which innovative activities take place.

In conclusion, this paper widens the concept of Schumpeterian patterns of innovation and moves it towards the broader concept of sectoral system of innovation (Malerba, 2002, 2004), in which not only firms, but also other actors involved in the innovation process (such as universities and research centres) are considered, and in which the knowledge base of sectors and technologies and the pervasiveness of innovations are crucial dimensions of the innovative process.

Acknowledgments

This paper originally draws from the project “The emergence of new industrial activities: fusing manufacturing and services” (TENIA)—contract no. SOE1-CT98-1119 funded by the European Commission, DG XII, Targeted Socio-Economic Research, IV RTD Framework Programme. The authors would like to acknowledge the financial support of the Italian Ministry for Education, Universities and Research (FIRB, Project RISC - RBNE039XKA: “Research and entrepreneurship in the knowledge-based economy: the effects on the competitiveness of Italy in the European Union”). We thank two anonymous referees, Ed Steinmueller, Nick Vonortas, Yannis Caloghirou, Nick Von Tunzelmann and participants to the Pavitt Conference and to seminars at Eindhoven Centre for Innovation Studies (ECIS) and CESPRI, for precious advice and suggestions. We also thank Gianluca Tarasconi and Andrea Pozzi for valuable research assistance.

Appendix A

Tables A1–A6.

Table A1
Patents and patenting firms in selected triples

Triple	Number of patenting firms	Number of patents (rank)
Printed circuit board	422	1018 (3)
Data processing system	362	1041 (2)
Mobile communication system	205	1735 (1)
Image processing system	202	942 (4)
Electrically insulating material	173	258 (17)
Data storage system	172	431 (9)
Light emitting element	170	372 (10)
Power supply unit	162	362 (11)
Central processing unit	161	359 (13)
Radio communications system	159	594 (8)

Table A2
Most important technological classes (four digits)

IPC code	Name
H04B	Transmission
H04N	Pictorial communication, e.g. television
H04L	Transmission of digital information, e.g. telegraphic communication
G06F	Electric digital data processing
H04Q	Selecting
G11B	Information storage based on relative movement between record carrier and transducer
H01L	Semiconductor devices; electric solid state devices not otherwise provided for
G01R	Measuring electric variables; measuring magnetic variables
G06K	Recognition of data; presentation of data; record carriers; handling record carriers
G01N	Investigating or analysing materials by determining their chemical or physical properties
H01J	Electric discharge tubes or discharge lamps
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
H04J	Multiplex communication
G11C	Static stores
G09G	Arrangements or circuits for control of indicating devices using static means to present variable information
H04H	Broadcast communication

Table A3
Most important technological classes (three 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

Italics: ICT technological classes according to Eurostat (2003).

Table A4
Most innovative firms

Firm	Patents	Triples	Specialisation (% of patents)
NEC (JP)	967	83	Mobile communication system (10.96%)
Matsushita Electronics (JP)	781	72	Information recording medium (13.83%)
Sony (JP)	747	65	Information recording medium (30.12%)
Siemens (DE)	601	73	Mobile communication system (17.80%)
Lucent Technologies (US)	559	73	Mobile communication system (28.44%)
Ericsson (SE)	507	61	Mobile communication system (26.43%)
Nokia Networks (FI)	409	40	Mobile communication system (50.12%)
Samsung Electronics (KR)	379	53	Mobile communication system (21.90%)
Philips (NL)	368	66	Image processing system (8.15%)
Canon (JP)	363	49	Image processing system (41.32%)
Toshiba (JP)	344	61	Image processing system (9.88%)
Motorola (US)	327	64	Mobile communication system (16.82%)
Texas Instruments (US)	318	55	Data processing system (12.26%)
IBM (US)	311	55	Data processing system (25.08%)
Fujitsu (JP)	299	51	Optical transmission system (10.03%)
Alcatel (FR)	290	52	Mobile communication system (19.66%)
Hitachi (JP)	287	60	Information recording medium (12.54%)
Nokia Mobile Phones (FI)	273	42	Mobile communication system (35.16%)
Sharp (JP)	271	51	Image processing system (14.76%)
Ericsson (US)	222	36	Mobile communication system (27.03%)

Table A5
T-test for equality of means between the two clusters (std. values)

Variables	*t*	Sig. (2-tailed)	Mean difference	Std. error difference	99% confidence interval of the difference	
					Lower	Upper
$ENTRY_{FIRM}$	9.136	0.000	1.469	0.161	1.151	1.788
$ENTRY_{PAT}$	9.731	0.000	1.526	0.157	1.215	1.837
GROWTH	3.006	0.005	0.906	0.301	0.291	1.520
PATENTS	0.082	0.934	0.018	0.215	−0.408	0.444
$HERF_{TECH}$	4.790	0.000	1.133	0.178	0.780	1.486
$CONC_{FIRM}$	2.171	0.032	0.156	0.072	0.014	0.299
UPCSOURCES	2.025	0.045	0.421	0.208	0.009	0.834
SELFSOURCES	−2.097	0.039	−0.265	0.126	−0.516	−0.014
$HERFSOURCES_{TECH}$	−3.585	0.000	−0.717	0.200	−1.114	−0.321
$HERFSOURCES_{FIRM}$	−1.048	0.297	−0.223	0.213	−0.645	0.198

Table A6
Non parametric test for equality of means between the two clusters

Variables	Kolmogorov–Smirnov *Z*	Exact sig. (2-tailed)
$ENTRY_{FIRM}$	3.382	0.000
$ENTRY_{PAT}$	3.499	0.000
GROWTH	1.693	0.006
PATENTS	0.764	0.604
$HERF_{TECH}$	2.519	0.000
$CONC_{FIRM}$	1.671	0.008
UPCSOURCES	1.027	0.242
SELFSOURCES	0.940	0.340
$HERFSOURCES_{TECH}$	1.780	0.004
$HERFSOURCES_{FIRM}$	1.089	0.186

References

Breschi, S., Lissoni, L., 2004. Knowledge networks from patent data. In: Moed, H.F., Glänzel, W., Schmoch, U. (Eds.), Handbook of Quantitative Science and Technology Research. Kluwer Academic Publishers, pp. 613–643.

Bresnahan, T.F., Trajtenberg, M., 1995. General purpose technologies: Engines of growth? Journal of Econometrics, Annals of Econometrics 65 (1), 83–108.

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

Corrocher, N., Malerba, F., Montobbio, F., 2003. How do new technologies emerge? A patent-based analysis of ICT-related new industrial activities. Innovation: Management, Policy & Practice 5 (2/3), 234–257.

Coulter, N., Monarch, I., 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), 1206–1223.

Courtial, J.P., 1994. A coword analysis of scientometrics. Scientometrics 31 (3), 251–260.

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

Dasgupta, P., Stiglitz, J., 1980. Industrial structure and the nature of innovative activity. Economic Journal 90 (358), 266–293.

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

Dosi, G., 1997. Opportunities incentives and the collective patterns oftechnological change. The Economic Journal 107, 1530–1547.

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

Fujimoto, M., Miyazaki, K., von Tunzelmann, N., 2000. Technological fusion and telemedicine in Japanese companies. Technovation 20, 169–187.

Granstrand, O., Patel, P., Pavitt, K., 1997. Multitechnology corporations: whey they have 'distributed' rather than 'distinctive core' competencies. California Management Review 39, 8–25.

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, A.B., Trajtenberg, M., Fogarty, M.S., 2000. The Meaning of Patent Citations: Report on the NBER/Case-Western reserve survey of Patentees. NBER Working Paper Series No. 7361.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localisation of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108, 577–598.

Kamien, M.I., Schwartz, N.L., 1982. Market Structure and Innovation. Cambridge University Press, Cambridge.

Klepper, S., 1996. Entry, exit, growth and innovation over the product life cycle. American Economic Review 86 (3), 562–583.

Kodama, F., 1992. Technology fusion and the new R&D. Harvard Business Review (July–August), 70–78.

Koumpis, K., Pavitt, K., 1999. Corporate activities in speech recognition and natural language: another 'new-science'-based technology. International Journal of Innovation Management 3 (3), 335–366.

Mahdi, S., Pavitt, K., 1997. Key national factors in the emergence of computational chemistry firms. International Journal of Innovation Management 1 (4), 355–386.

Malerba, F., 2002. Sectoral systems of innovation and production. Research Policy 31 (2), 247–264.

Malerba, F. (Ed.), 2004. Sectoral Systems of Innovation. Cambridge University Press, Cambridge.

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 13 (4), 411–434.

Malerba, F., Orsenigo, L., 1995. Schumpeterian patterns of innovation. Cambridge Journal of Economics 19, 47–65.

Malerba, F., Orsenigo, L., 1996. Schumpeterian patterns of innovation are technology specific. Research Policy 25, 451–478.

Malerba, F., Orsenigo, L., 1997. Technological regimes and sectoral patterns of innovative activities. Industrial and Corporate Change 6 (1), 83–117.

Miyazaki, K., 1994. Search, learning and accumulation of technological competences: the case of optoelectronics. Industrial and Corporate Change 3 (3), 631–654.

Montobbio, F., 2003. Sectoral patterns of innovations and market share dynamics. Cambridge Journal of Economics 27, 523–545.

Montobbio, F., 2004. Sectoral dynamics and structural change. stylized facts and 'system of innovation' approaches. In: Malerba (Ed.), Sectoral Systems of Innovation. Cambridge University Press, pp. 42–69.

Mowery, D.C., Nelson, R.R. (Eds.), 1999. The Sources of Industrial Leadership. Cambridge University Press, Cambridge.

Nelson, R.R., Winter, S., 1982. An evolutionary Theory of Economic Change. The Belknap Press, Cambridge (MA).

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

Noyons, E.C.M., 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), 68–81.

OECD, 1994. Using Patent Data as Science and Technology Indicators—Patent Manual. OECD, Paris.

OECD, 2002. Measuring the Information Economy 2002. OECD, Paris, http://www.oecd.org/dataoecd/16/14/1835738.pdf.

Pavitt, K., 1984. Sectoral patterns of technical change: towards a taxonomy and theory. Research Policy 13, 343–373.

Pavitt, K., 1994. Key characteristics of large innovating firms. In: Dodgson, M., Rothwell, R. (Eds.), The Handbook of Industrial Innovation. Elgar Publishing, Cheltenham, UK.

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

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

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.B., 2002. University versus corporate patents: a window on the basicness of invention. In: Jaffe, A., Trajtenberg, M. (Eds.), Patents, Citations & Innovations. MIT Press, Cambridge, pp. 51–87.

Tushman, M.L., Anderson, P., 1986. Technological discontinuities and organisational environment. Administrative Science Quarterly 31, 439–465.

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

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

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.