# Network Analysis for Patent Citation Data: "H04Q-007" - Deconstructing the Technological Blob of Mobile Telecommunication

***Harald Pier***
*Swiss Federal Institute of Intellectual Property*
*Einsteinstrasse 2*
*CH-3003 Bern*
*Harald.Pier@ipi.ch*

***Katja Rost***
*University of Bern*
*Institute of Innovation Management*
*Engehaldenstrasse 4*
*CH-3012 Bern*
*Katja.Rost@iim.unibe.ch*

***Thorsten Teichert***
*University of Bern*
*Institute of Innovation Management*
*Engehaldenstrasse 4*
*CH-3012 Bern*
*Teichert@iim.unibe.ch*

***Iwan von Wartburg***
*University of Bern*
*Institute of Innovation Management*
*Engehaldenstrasse 4*
*CH-3012 Bern*
*Iwan.vonWartburg@iim.unibe.ch*

## Network analysis for patent citation data:

### "H04Q-007" - Deconstructing the Technological Blob of Mobile Telecommunication

# 1 Revealing the patterned structure of patent data

The incentives for firms, regions or nations to compare their competitive position are straightforward, especially because it is widely assumed that technological change is a "good" phenomenon. From a theoretical point of view however studying technological change using patent indicators has not been considered important before the advent of the so-called "New Growth Theory":

> *"[…] until recently, economists have tended to view technology as a black box that affected the economic system but that was itself driven largely by exogeneous noneconomic forces, such as the advance of science." (Jaffe 1998, p. 8)*

Given that the study of technological change is an important research field (Rosenberg 1976, 1982), the question remains of how such a complex, interrelated and cumulative process like technological change might be analyzed. Because of the complexity involved, indicators of technological change need to reflect temporal and causal "path dependencies" (Arthur 1989) indicating cumulative technical learning (Cohen and Levinthal 1989, 1990, 1994): they must help explaining what the impact of the economic agents' prior investments on the marginal outcome of subsequent investments made by others is. This is however all the more difficult because:

> *"Technological change is driven by an investment process that produces a form of capital that is hard to see or measure." (Jaffe 1998, p. 8)*

Patent documents[1] include information that is advantageous for studying technological change and economic growth. Patents are a direct output category of industrial R&D and other inventive activity. They mirror the cumulative process of technological change: on the one hand patent data enable longitudinal research and on the other they contain citation information that link different patents at different stages of technological development. They cover almost every field of technology that is useful for analyzing the diffusion and the development of key technologies. Patent data provide additional benefits such as global

---

[1] "A patent is a document, issued by an authorised governmental agency, granting the right to exclude anyone else from the production or use of a specific new device, apparatus, or process for a stated number of years" (Griliches 1990, p. 1662). For an overview about the patent application, examination, granting and enforcing process see Harhoff and Reitzig (2001).

geographic coverage and accessibility through large commercial and free electronic databases. They contain standardized details of interest like information about inventors and assignees, years of inventions, claims covered and the like. They are classified in patent classification schemes that allow studying on different levels of aggregation.

There are several drawbacks for patent data analysis that have important implications for the validity of patents as measures of the competitiveness of firms, regions or nations.[2] Firstly and most importantly, not all inventions are patented.[3] Secondly, invention is not equal to innovation.[4] One could of course argue that, by patent law definition, a granted patent must be potentially commercially exploitable. Therefore, one could further argue that patent data contain information not only about invention but also about innovation.[5]

Empirical studies indicate that patent data can be indicators for the study of technological change in many industries. Arundel and Kabla (1998), extending the work of Scherer (1983) and Mansfield (1986), studied propensity rates for European firms in different industries. They define the propensity rate as the share of inventions that lead to a patent application[6]. They found that the average propensity rate for product innovations is 35.9 %, with values

---

[2] To review different caveats for using patents as a proxy to study technological change, see for example Pavitt (1985), Basberg (1987), Griliches (1990), Archibugi (1992), Verspagen (2000).

[3] "First, not all inventions meet the patentability criteria […] (the invention has to be novel, non-trivial, and has to have commercial application). Second, the inventor has to make a strategic decision to patent, as opposed to rely on secrecy or other means of appropriability. Unfortunately, we have very little idea of the extent to which patents are representative of the wider universe of inventions, since there is no systematic data about inventions that are not patented. This is an important, wide-open area for future research. (Hall et al., 2001, p. 5)

[4] Grupp (1998) discusses the validity range of patent data by further dividing patent applications that meet the patentability criteria into two groups called "innovation relevant inventions" and "applications without economic relevance". By doing so, he explicitly refers to the difference between "invention" and "innovation". Whereas 'invention' refers to a supply side output of inventive activity that may or may not become economically relevant, 'innovation' takes the view of the demand side and refers to inventions that were introduced in the marketplace.

[5] Griliches (1990, p. 1669) argues in this direction when he states that "[…] a patent does represent a minimal quantum on invention that has passed both the scrutiny of the patent office […] and the test of the investment of effort and resources by the inventor and his organization […], indicating thereby the presence of a non-negligible expectation as to its ultimate utility and marketability."

[6] Scherer discusses the "propensity to patent" measure as the number of patents per unit of expenditure on R&D (Scherer 1983). Mansfield (1986) uses the percentage of patentable inventions that is patented as the propensity rate.

ranging between 8.1% for textiles and 79.2 % for pharmaceuticals followed by chemicals (57.3%).

Patent data have been used as economic indicators for many different research questions (Schmokler 1966, Scherer 1982, Grupp 1998). Comparative studies on micro- (individual inventors and firms), meso- (firm networks and regions) and macro-levels (technologies and nations) discuss

> *"Questions about sources of economic growth, the rate of technological change, the competitive position of different firms and countries, the dynamism of alternative industrial structures and arrangements all tend to revolve around notions of differential inventiveness: What has happened to the "underlying" rate of technical and scientific progress? How has it changed over time and across industries and national boundaries?" (Griliches 1990, p. 1661)*

Thus, no matter how large the range of specific research interests might be, researchers today rely more than ever on patent data as a rich, structured and standardized data repository that reveals the patterned structure of technological change.[7]

## 1.1 State-of-the-art of patent analysis

### 1.1.1 Simple patent counts based on patent classification systems

The most common method in early patent data studies was simply to count them and to compare how many patents (applications and granted patents) had been assigned to different entities. These studies are called simple patent counts (SPC).

SPC observe patenting trends on different levels such as inventors, firms or firm networks, fields of technology, regions or nations. Because patent data contain historical information, it is also possible to count patent distribution changes over time (OECD 1994). Well known measures for "benchmarking" the inventive outcome of competitive entities are specialization indices like the "Revealed Technology Advantage" (RTA) and the "Revealed Patent Advantage" (RPA).

---

[7] Griliches (1990, p. 1661) makes a somewhat more pessimistic guess about research's learning curve concerning patent data analysis. He states that "the idea that something interesting can be learned from such data tends to be rediscovered in each generation."

The RTA is the ratio of a nation's patent share percentage in a specific technological subdomain divided by the nation's patent share percentage in the whole technical field (Soete and Wyatt 1983). RTA values range from zero to infinite. There may be a bias when a nation only accounts for a very small share of patents in the whole technical field because the denominator becomes too small. The RPA is an adjusted RTA measure and is defined by: 100 ln RTA. Both RTA and RPA are measured over time to evaluate the persistence of competitive (dis-)advantages of nations.

These measures have been widely used by technical non-experts to derive political policy implications. Thus, it is of paramount importance that the underlying categorization of technology, the classification system, is carefully constructed and maintained. This leads us to a first problem of SPC: the *classification* problem.

The *classification problem* results from the fact that patents are classified in Patent Classification Systems (PCS).[8] These systems emerged as physical storage and retrieval utilities that helped patent examiners storing and locating prior art information. They have evolved into highly complex trees of technology with many branches and even more sub-branches. The classification problem addresses the necessity and difficulty of mapping the technological categories included in a PCS to economically meaningful entities like products, firms or industries. Early SPC studies did not make this effort:[9]

> *"[…] with one notable exception […], almost all attempts to relate patent numbers to industrial data use the subclass system as their basic unit of assignment" (Griliches 1990, p. 1666).*

Grupp (1998) remarks thoughtfully that any mapping table, no matter how well considered and carefully constructed, can become obsolete in cases of radical innovations combining new technologies with new market applications. Thus, useful mapping tables can be constructed for a case-by case analysis but not for a universal purpose.

[8] For a discussion of the International Patent Classification as a patent classification system see section 1.2 below.

[9] For an overview of the diverse mapping efforts and the involved difficulties see Griliches (1990), Trajtenberg (1990), and Debackere (2002).

To illustrate the inherent ambiguity in the task of constructing a mapping table or concordance between technology fields in PCS and for example industries, one may think about the different assignment criteria that can be chosen from (Griliches 1990). A patent may be assigned to the industry where it was invented (origin assignment). It may be assigned to the industry where the invention will be produced (assignment to production). It may also be assigned to the industry where the invention will be used (assignment to use).

A second problem with SPC studies is the *patent value* problem that resides in the accepted observation that the distribution of patent values is highly skewed: a small fraction of a patent's sample mostly accounts for the largest part of its value (Griliches 1990, Harhoff et al. 2002). A simple count of the number of patents equal these differences in value and importance out.[10] Debackere (2002) enumerates the different avenues for dealing with the patent value problem that have been followed by researchers like for example:

- The use of geographical coverage of a patent as a value indicator measured by the number of "designated states" for a EPO or WO patent or by the number of patent family members for national patents (Mogee and Kolar 1994)
- The use of patent renewal information measured by the lifetime of patent protection and derived patent mortality rates (Pakes and Schankerman 1984, Pakes 1986, Lanjouw et al. 1998)
- The breadth of a patent measured as the number of technology classes assigned to a patent during the examination procedure (Lerner 1994, Harhoff et al. 2002)[11]

Having mentioned all these difficulties, how can one judge what can really be measured with patent statistics? As evidence suggests, simple patent counts are closely associated with the input side of the inventive process, measured mainly with R&D expenditures (Schmokler 1966, Trajtenberg 1990, Griliches 1990). And yet they did not prove useful to explain output indicators like the performance or value of firms (Trajtenberg 1990).

---

[10] Therefore it is only adequate to depend on patent numbers if the analyzed sample is large enough to assume an identical value distribution. Usually, this is the case when comparing the technological strength of nations in certain technological fields.

[11] Harhoff et al. (2002) did not find evidence for a causal relation between the number of four digit IPC classes assigned to a patent and its value.

Therefore, in SPC studies the phenomenon analyzed might at best be called "invention". What can be counted is which firm, region or nation accounts for the most inventions and how these numbers relate to the investments made. For SPC studies it is not legitimate to call the phenomenon studied "innovation", since innovation in almost every definition requires a market introduction. This requirement is only fulfilled by a fraction of the patents counted.

> *"The dream of getting hold of an output indicator of inventive activity is one of the strong motivating forces for economic research in this area" (Griliches 1990, p. 1669).*

By relying on SPC this dream will stay an unfulfilled one: to become an indicator of *innovativeness*, patents need to be commercialized.[12] Patent data analysis will never be able to measure *directly* the amount of financial and intangible returns to an invention. Thus, to be allowed to continue to dream, another way of dealing with the discussed problems of SPC has been followed since around the mid-eighties: "patent citation analysis".

### 1.1.2 Patent citation analysis

*Patent citation analysis* has been receiving increasing levels of attention in recent years (Michel and Bertels 2001, Pilkington et al. 2002). This leap in attention has been enabled by improved patent data quality and – foremost - easier as well as ubiquitous access through new information technologies like the internet.[13]

Patent citation analysis is based on the examination of the citation links among different patents and between patents and scientific literature. When applying for a patent, the assignee has to prove the novelty, non-obviousness and usefulness of his invention. For this reason, his own invention is compared with prior art both by the inventor and the patent examiner.[14]

---

[12] An even stronger requirement would be that the inventors were also able to appropriate a substantive part of the returns to an innovation (Teece 1986).

[13] Before the availability of electronic patent document databases containing bibliometrical information it was not possible for researchers to perform patent citation analysis with a certain ease. Thus, this technique for measuring the value of patents coevolved with the enabling information technology (Schwander 2000).

[14] The incentive for an inventor to explicitly cite all prior art that is known to him and that may be a danger for one of the claims contained in his patent can be either the anticipated ease to defend the patent in a possible litigation process or the legal requirement to do so. The legal requirements vary internationally. US patent law for example requires an inventor to disclose in an information disclosure statement all prior art that is relevant to patentability of the invention in question ("duty of candour"). Failing to keep up to this duty can cause severe penalties (Akers 2000). However, there is no such duty in Europe. Therefore, to avoid biases because of different

*“Patent data include citations to previous patents and to the scientific literature. These citations open up the possibility of tracing multiple linkages between inventions, inventors, scientists, firms, locations, etc. In particular, patent citations allow one to study spillovers, and to create indicators of the "importance" of individual patents, thus introducing a way of capturing the enormous heterogeneity in the “value” of patents.” (Hall et al. 2001, p. 4)*

A patent can be valuable from a technological and/or from an economical point of view. The long-term *technological* value is the importance of a patent as a foundation for subsequent technological inventions. It can be approximated by the number of times a patent is cited. This relation is validated by a number of evidence (Carpenter et al. 1981, Albert et al. 1991, Harhoff et al. 2002). The *economic* value of patents is measurable by their impact on output success measures of the entity studied. For a firm, the question is how strong a relation between the citation information of patents and the performance of the firm can be found. A positive relationship between “times cited” of patents and firm performance has been found and therefore one may conclude that more often cited patents seem to have a higher economic value (Trajtenberg 1990, Griliches 1990, Breitzman and Thomas 2002, Harhoff et al. 2002). Furthermore, there is evidence that firms with highly cited (forward citations) and highly citing (backward citations) patents achieve better stock market valuations (Deng et al. 1999) as well as sales and profits (Narin et al. 1987).[15]

Thus, by shifting the focus away from counting patents to citation information contained in patent data, output success may be measured *indirectly.* Patent citation analysis legitimates researchers to call the phenomenon studied not only “invention” but also “innovation”.

Figure 1summarizes the methodological approaches to patent statistics discussed so far:

---

average amounts of citations per patent caused either by these different disclosure requirements or by different examination procedures, we only use European patents in our sample of the H04Q-007 data.

[15] Narin et al. (1987) study the pharmaceutical industry, where patenting and intellectual property is significantly more important than in other industries.

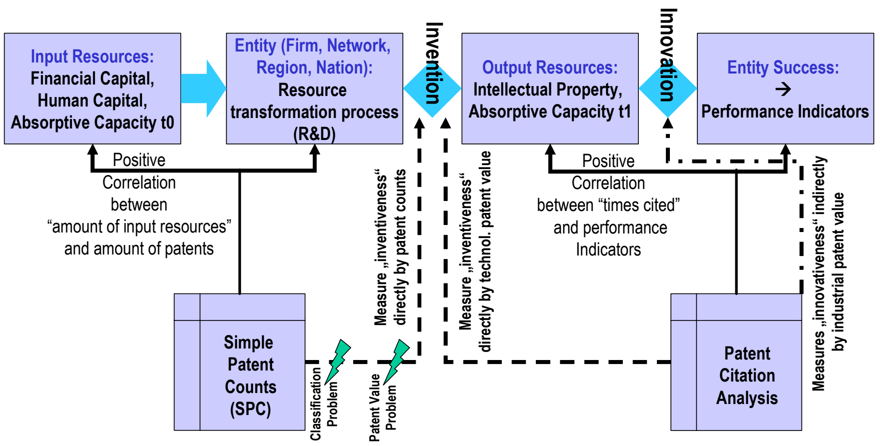



**Figure 1: Invention, innovation, simple patent counts and patent citation analysis.**

Patent citation analysis bases on either "Backward" measures (derived from the citations made by a patent) or "forward" measures (derived from the citations that a patent subsequently receives from other patents).[16] As discussed above, forward citation studies have used counts of patent citations as a measure of its quality in terms of technological or economic value (Harhoff et al. 1999, Henderson et al. 1998, Trajtenberg 1990). Studies using backward citation information can be classified by three dimensions. First, some studies investigated spillovers called "knowledge flows" between technology classes (Rosenkopf and Nerkar 2001) or geographic regions (Jaffe et al. 1993, Tijssen 2001, Verspagen 2000). Then, other studies used backward orientated citations as a means to partition the technology space and to understand the nature of the relationships between technologies (Stuart and Podolny 1996) or between technologies and firms (Pilkington et al 2002). Finally, backward citations were used to measure the interdependence between particular technologies (Fleming and Sorenson 2001).

For studies both of forward and backward citation data, a variety of measures has been proposed (Trajtenberg et al. 1997). For example, the already introduced RTA measure was

[16] Of course this separation is an artificial one because all citations stem form documents that cite. Hall et al. (2001, p. 7) call this fact the "inversion problem": "The inversion problem refers to the fact that the original data on citations come in the form of citations made (i.e. each patent lists references to previous patents), whereas for many of the uses (certainly for assessing the importance of patents) one needs data on citations received. The trouble is that in order to obtain the citations received by any one patent granted in year t, one needs to search the references made by all patents granted after year t. Thus, any study using citations received, however small the sample of patents is, requires in fact access to the whole citations data, in a way that permits efficient search and extraction of citations."

reincarnated by using citation counts instead of patent counts and by assigning a new label to this measure: the attractivity index.[17] Jaffe summarizes these measures by dividing them into

> *"[…]three categories: importance measures are based on the number of citations made or received; distance measures relate to the proximity or remoteness of the cited or citing patents, across both time and technology space; and originality or generality measures relate to the dispersion of citations made or received across different areas of technology space. We also examined the extent to which the citations made by patents were to scientific articles rather than to other patents as an indicator of the closeness of the invention to basic science. […] We also proposed that the fraction of "self-citations" - citations that come from patents assigned to the same organization - was an indicator of the originating organization's successful appropriation of the subsequent fruits of that research. (Jaffe 1998, p. 8)"*

In Figure 2 we try to visualize the state-of-the-art of patent citation analysis studies. It is important to state that the studies so far only used single-stage citation information. They investigated citation links between technologies, between firms, between firms and technologies, between nations, and between technology fields and science. What they failed to do was to investigate longer citation chains and getting more out of the historical citation information. In this paper, by building a relational database management system and relying on network measures from social network analysis, we advance the state-of-the-art in this direction.

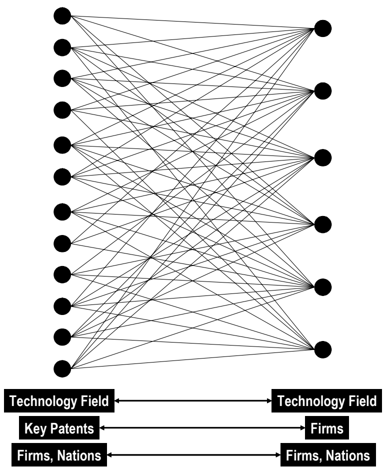



**Figure 2: Cross-citation studies**

[17] CHI Research introduced the "citation performance index" to compare the 10% most highly cited patents of a country with those of the world (Debackere 2002).

## 1.2 A closer look at the information sources

### 1.2.1 IPC

In this paragraph, the characteristics of the International Patent Classification and the resulting limitations for purposes of patent analysis in the context of technological and economic studies are discussed. To begin with, a few words about the origins of patent classification will help clarify these limitations.

#### *1.2.1.1 Historical evolution*

Traditionally, patent documents were stored in "shoebox"-like storage elements that occupied some kilometers of shelf space in patent documentation centers and patent offices. This constituted the technical literature defining the state-of-the-art used by patent examiners to compare against new inventions. In order to be able to locate a given technology in this paper collection, it was necessary to have a retrieval tool capable of assigning one single location to every document. Every patent office around the globe therefore had its own classification scheme with this common requirement: as long as patent practitioners dealt with paper archives, the document had to be archived at one specific location.

Even before the advent of electronic databases, enhanced communication between national patent offices necessitated common guidelines for the storage and retrieval of technical documents. This gave birth to the International Patent Classification (IPC), which is based on the Strasbourg Agreement Concerning the IPC, concluded in 1971 and introduced in1975. Soon, the technological evolution resulted in the need to update the initial edition of the IPC. Since then, a new edition of the IPC has been issued every 5 years, so that currently the 7$^{th}$ edition of the IPC is in place. The 7$^{th}$ edition of the International Patent Classification (IPC) and the European Patent Classification (EC) currently divide technology into eight sections with 61560/ 119617 subgroups each.

Today, the industrial property offices of more than 90 States, four regional offices and the International Bureau of WIPO under the Patent Cooperation Treaty (PCT) actually use the IPC. This means that patent examiners of these offices apply the IPC guidelines[18] when

[18] http://www.wipo.org/classifications/en/ipc/manual/index.htm

classifying ng patent applications, so that they can be retrieved by their colleagues wishing to search for a specific technology. Being conceived primarily as a documentation and search tool, the IPC is still focused on a paper-equivalent type of archiving, implying that documents should essentially bear a single classification code. Generally the classification occurs by function and not by application, which has been considered as less ambiguous. There are however exceptions and in order to direct patent practitioners towards the appropriate set of documents for their search request, a set of rules and additional information is required.

Together with advanced electronic possibilities of classification, the IPC has reached a point where a serious revision is necessary. Exactly this effort is on its way and in a few years from now we will see a modernized system split into a core and an electronically published advanced level that is better suited to future technological advances.

#### *1.2.1.2 Consequences for patent analysis*

Based on the fact that the IPC is the true worldwide classification system for technical information, what may the implications of this historical evolution of patent classification for purposes that are beyond storage and retrieval of technical information be?

The refined categorization should enable a technical non expert to monitor developments in different technical fields by comparing patent activity in different subgroups. Being conceived as a storage and retrieval tool however the IPC has some inherent shortcomings when applied for patent analysis of any kind.

Firstly, the IPC/EC schemes have evolved historically, so the resulting perspective on technology and its categories does not necessarily fit with the perspective adopted by a particular company or researcher. Secondly, the IPC, currently being updated every 5 years, cannot cope with the pace of technical change in certain industries such as mobile telecommunication. Furthermore, while the classification was updated every five years, the classification on the patent documents itself has not changed accordingly, so that statistics over a longer period of time have to be carefully verified. The more dynamic and precise EC is derived from the IPC and not available on all documents.

Technologies such as electronics, wireless communication, but also advances in the mastering of materials and ever smaller structures have almost ousted the strict application of the IPC philosophy. Their application is almost by default multidisciplinary. As a result, several classification codes on a patent document today are more a rule rather than the exception. Thus, the statistical weight of patent documents can vary in proportion to the number of different classification codes found on the document.

Moreover, not all examiners worldwide have the same acquaintance with the application of the classification guidelines and even if they have, certain "cultures" of classification have evolved that differ from one country to another. We have also seen that the IPC classifies by function, while the economy is more often interested in applications, technology food chains[19], relationships between competitors, or regulatory interactions with markets.

### 1.2.2 Citation

Patent citation measures strongly depend on the rules that examiners follow when adding citation data to a search report and during examination. It is therefore essential for researchers relying on patent data analysis to be coherent with the European examiners' practice. For this purpose, it is worthwhile taking a look at the guidelines and practices for patent examination at the EPO, in particular the search for prior art:

> *"A key weapon in the armoury of a patent information specialist is an understanding of how these document citations arise and, perhaps more importantly, their relevance to the validity of the European patent in question and any corresponding patents or applications." (Akers 2000, p. 309)*

#### *1.2.2.1 The Search procedure at the European Patent Office*

Before the implementation of the BEST[20]-Project, search and substantial examination were strictly separated tasks at the EPO. The search was carried out by dedicated search examiners at The Hague[21], whereas the substantial examination was done by a colleague in Munich. The basic idea behind this split approach was that the search examiner should be able to fully concentrate on the search task, while the person doing the substantial examination should not

---

[19] Trajtenberg refers to 'lines of innovation' (Trajtenberg 1990, p. 3).

[20] **B**ringing **E**xamination and **S**earch **T**ogether, an approach to unify search and substantial examination in order to increase efficiency and make use of synergies. It is now becoming the standard for all newly hired examiners to perform both, search and substantive examination.

[21] Searches for European patents can also be carried out at "authorized" national offices, particularly when the language of the patent application is not one of the official languages of the EPO.

be influenced in his judgement of newness and presence of an inventive step by all the prior art that would otherwise have passed before his eyes. [22]

The task of the search examiner is to identify the closest state-of-the-art patent to the claims of the patent application. The search documentation, both in paper and electronically, is at the examiner's disposal and enables him to carry out a thorough an assessment as possible. The documentation includes the PCT minimum documentation, which is a well-defined collection of patent documents for international patent applications and thus represents a lower limit for the search standard.

At several places in the guidelines, the examiner is asked to "exercise his judgment". Certainly, there is no proof of completeness of a search report, which puts more emphasis on the training and experience of the examiners. The European patent office claims to have relatively low miss rates[23]. What documents are consequently cited in the European search reports?

#### *1.2.2.2 Content of the search report*

As mentioned above, the rules concerning the prior art to be added to the search report will strongly influence our measure of proximity. Therefore, we will now investigate what type of subject will be added to the search reports.

The examiner is asked to base his search on the claims, usage description and drawings for interpretation where necessary. At first, the focus is on locating novelty-destroying material, thus subject matter pertaining to the **same** technological field is searched for. Only if no such documents are found will the search be extended to **similar** or **related** fields. The search is to be stopped when documents have been found clearly demonstrating a lack of novelty in the entire subject-matter of the claimed invention. If this is not the case, the examiner stops his search when the probability of discovering further relevant prior art becomes very low.

---

[22] The novelty and inventive step judgement should take the inventor's perspective, and not be a posteriori view of the invention in question.

[23] This can be confirmed by additional, much more profound searches that are carried out to identify novelty destroying subject matter. The Swiss Federal Institute of Intellectual Property regularly performs such searches.

The selection of documents is guided by the following principles: [24]

- Most relevant documents first.
- Less relevant documents only when they concern aspects or details of the claimed invention **not found in the documents already selected for citation**.
- In borderline cases or cases of doubt a larger amount of documents should be cited.
- However, no more documents than **necessary** are to be cited; if there are several documents of equal relevance, the search report should only cite one of them
- The cited documents should be preferably in the language of the application if available (e.g. patent from the same family).

At the European Patent Office, examination experience has shown that most of the relevant information on the criteria of patentability is obtained from 1-2 documents:

> *"According to the EPO philosophy a good search report contains all the technically relevant information within a minimum number of citations." (Michel and Bettels 2001, p. 189)*

#### *1.2.2.3 Consequences for patent analysis*

Compared to simple activity counts of patent activity in respective technical fields, patent citation data are advantageous in terms of validity and reliability:

- Patent citation data may have a high validity in terms of content quality attributed. To receive a granted patent, the inventor may carefully explore the state of the "prior art". Foremost, an examiner will evaluate both the invention as well as the scope of applied claims by performing an in-depth review of existing patents and non-patent literature and documents his findings in the citations[25]. Therefore, patent citation data can be interpreted as especially rich historical information as well as a proxy for information spillovers (Jaffe et al. 1993, Almeida 1996).

---

[24] Guidelines for examination in the EPO, Part B: Guidelines for the Search, Chapter IV: SEARCH PROCEDURE AND STRATEGY, 3. Procedure after Searching

[25] For an elaborated overview of the patenting process see Harhoff and Reitzig (2001).

- Citation data have rich cognitive contents and are thus of high reliability: Whereas it takes on average 10-30 minutes to classify a patent application, a search for state-of-the-art and its evaluation may well take up to three to four days, and profits from all preceding classification work[26].

Given that the advantages of citation analysis are now obvious, some considerations about the applicability of generic methodology for citation analysis to patent information shall be made. Citation analysis has its origins within bibliometric studies of scientific authors' or academic journals' citation behaviour. By analysing patent citation information, it has to be taken into account that there are important differences. These differences are rooted in the search procedure as described above. Most importantly, because of the minimum number of references that is targeted, cited references point only to the 'closest' prior art. Consequently – and in contrast to citations found in literature - citation entries from single documents are unable to bring together whole networks of technological inventions.

# 2 Patent citation network analysis

## 2.1 Network analysis for evaluating patent citation data

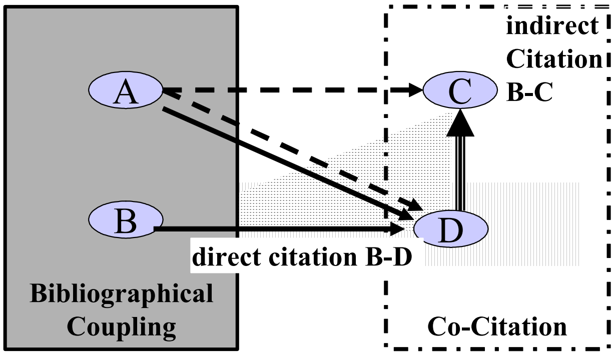
A network for patent analysis consists of nodes (patent documents) and linkages (edges) between them (citing and cited citation information). Connections represent flows between the nodes. Two nodes i and j are connected if some information flows between the two documents. A network consisting of patent families is completely connected if every node can be linked to every other node through a number of intermediary nodes. The smallest number of edges between two nodes is called the "geodesic path". The importance of patent families can be described in terms of their network centrality, their power or their connectedness[27].

Bibliometrical analysis draws on different information types. Firstly, one can separate 'direct' from 'indirect citations' and secondly, 'bibliographical coupling' from 'co-citations' (Figure 3: Citation types and adjacency matrix. If a patent cites or is cited by another patent, this link is called a *'direct citation'* link (B→ D). If a cited patent in turn cites another patent, a second direct citation link (D→ C) is established. Furthermore, an *'indirect citation'* chain between the first citing and the last cited patent is in place (B→D→C). A *'bibliographical coupling'* refers to a citation link where more than

[26] See: http://www.european-patent-office.org/epo/pat_examiner.htm.

[27] See Wassermann and Faust (1994).

one patent cites the same patent, i.e. if a patent is cited by different patents (A/B→D). A *‘co-citation link’* occurs if patents are cited in the same patent (A→C/D).




***Adjacency Matrix***

| | A | B | **C** | **D** |
|---|---|---|---|---|
| **A** | 0 | 0 | 1 | 1 |
| **B** | 0 | 0 | 0 | 1 |
| C | 0 | 0 | 0 | 0 |
| **D** | 0 | 0 | 1 | 0 |

**Figure 3: Citation types and adjacency matrix**

Traditionally, patent analyses take one or a few patent documents as anchors to start hunting for direct references. Such an approach tends to reveal more co-citation groups of earlier patents than bibliographical “couples”. It disregards the actual power of patent citation data, i.e. to identify and analyze information spillovers between networks of patents. This is especially harmful, as EPO-examiners do not list all related documents in the search report but carefully select them. Based on European guidelines, primarily closely related patents are found in search reports. Consequently – and in contrast to citations found in literature - citation entries regularly miss reference to basic inventions of a technology field. To mirror actual developments in a certain technical field, citation analysis should rely on everything, bibliographical coupling, co-citations, direct and indirect citations. Using hierarchical cluster analysis based on citation information that is collected within a relational database one can ‘capitalize’ on all of the citation types mentioned above.

## 2.2 Methodological considerations for patent citation network analysis

Out of the various measures which network analysis provides, the reachability between nodes is of central importance to us. In the context of patent data, a refinement of the basic model of reachability is required. Firstly, the equal treatment of any two direct connections between two actors is to be questioned. Secondly, the information about indirect linkages is to be examined. Finally, the handling of biographical coupling is investigated.

### 2.2.1 Reachability and Proximity

Reachability is based on the concept of geodesic paths. It is achieved by assessing the shortest possible connection between any two nodes, both direct as well as indirect: In the simplest case, all direct linkages of a node are summarized in an unweighted measure of connectedness. At the same time, indirect linkages are generally assumed to be of smaller value. Both direct and indirect measures are added to assess the proximity of two nodes. Intuitively, this makes sense in the sphere of social relationships. Given that one does not have any other a-priori information available, a simple count treats the direct linkages as being of equal and compensatory value. Indirect linkages are then treated as additional but less strong linkages, as they require some sort of involvement of the nodes between. This is easily understood if one thinks about one's own personal network and its usage, be it either for personal or business purposes.

The meaning of proximity and thus its operationalization and measurement clearly depend on the specific objectives of an investigation. In our context of patent data analysis, citation information is to be used for classifying technologies. Here, we are interested in clustering those patents which relate to the same technology field or area. Accordingly, proximity of any two patents is interpreted as their technological relatedness, i.e. the degree with which both patents build on identical technological principles. At this stage, we do not pursue a differentiation based on location along the path of technological development, thus we treat patents both in the early or late phase of technological development as identical.

### 2.2.2 Measurement of direct citation links

Equal weights of direct citation links can be regarded as a reasonable assumption in a non-informative a-priori situation; however weights may be attributed to the interactive patterns of nodes based on information about their relative distances, costs, strengths or probability values. In this regard, patent citation data provide specific clues on the strengths of relationships even to the technical non-expert. We argue that the more patents are referenced, the broader the technological base of the citing patent is. This holds, since patent examiners are required to limit the cited references to the least amount needed. If a *single* prior patent encompasses all relevant state-of-the-art, only this one will be referenced. In this case, the citing patent can be regarded as an immediate successor of the former one. Accordingly, the proximity between those patents is of maximum value. If *two* patents are cited, the new invention can be assumed to base equally on both prior patents. Vice versa, the proximity between the citing and any of the two cited patents should be regarded as only half of the maximum value. This holds, since the new invention is likely to integrate certain aspects of both former ones and thus can be regarded as a hybrid development.

Citation patterns are likely to differ across technology fields (Hall et al. 2001). Thus, measures of proximity ought to be standardized in order to allow cross-patent comparisons. For our study, we divide the number of direct citation links between two patent families by the total number of references stemming from the citing patent. This scheme serves as a standardization, by which single relationship strengths are treated in a probabilistic manner: The weighting scheme works as follows. If a citing patent references only a single other patent, a probability value of 1 is attributed to this relationship. For any other case, a probability of 1/n is allocated to the relationship between two patents, whereby the citing patent references a total of n other patents. This probabilistic interpretation limits the strength of direct relationships between any two patents to a maximum value of one.

### 2.2.3 Measurement of indirect citation links

Whereas indirect links are in general of less value than direct links, this does not hold true in case of patent citations. As explained above, patent citations differ from article citations in regard to the fact that the references are strictly limited to the nearest possible patent. It is however reasonable to assume that the technological foundation of citing patents does not only encompass the most recent developments cited directly. It also draws on basic principles provided by earlier patents. Connections to basic patents are revealed by indirect linkages which stem from the flow of citation. Thus, the commonly applied attenuation of linkage value should not be applied. In contrary, we argue that no attenuation factor is required at all if one is interested in the cumulative development of technologies.

Given that a patent A cites exclusively patent B which in turn solely cites another patent C, a unique development path can be assumed which stems from C and leads to A. Since both A and B ultimately originate from C, both can be regarded as technological improvements of the latter. Thus, we have to conclude that A does not only base on B but in the same way on C. Accordingly, we attribute an identical proximity of A versus B, B versus C and A versus C.

A differentiated view is necessary if more than one patent is cited at any one level of the citation flow. Here, the reduced proximity between those directly linked patents influences the entire chain of relationships. Accordingly, the indirect relationship between any two patents X and Y is assessed by multiplying the strengths of the direct relationships connecting both patents.

### 2.2.4 Dealing with directional changes: Bibliographical coupling

Traditionally, network relationships are either treated as having no direction or as a vector. In the latter case, only a one-directional influence is analyzed. We argue that for analyzing patent citation information, a differentiated view is necessary.

First of all, we are interested in proximity between any two patents as a symmetrical measure: Patent citation data are however of inherent asymmetric nature. This is due to the time-based character of citation information, whereby only the younger patent is able to cite the elder one. Evidently, this asymmetry does not hold true from a contextual perspective. Thus, proximity measures should be treated as identical in both directions of the citation flow. Indirect linkages should be valued analogously: Symmetry holds as long as the citation flow is unidirectional. In this case, the distance between any two patents is calculated as the product of the values of their intermediating direct relationships.

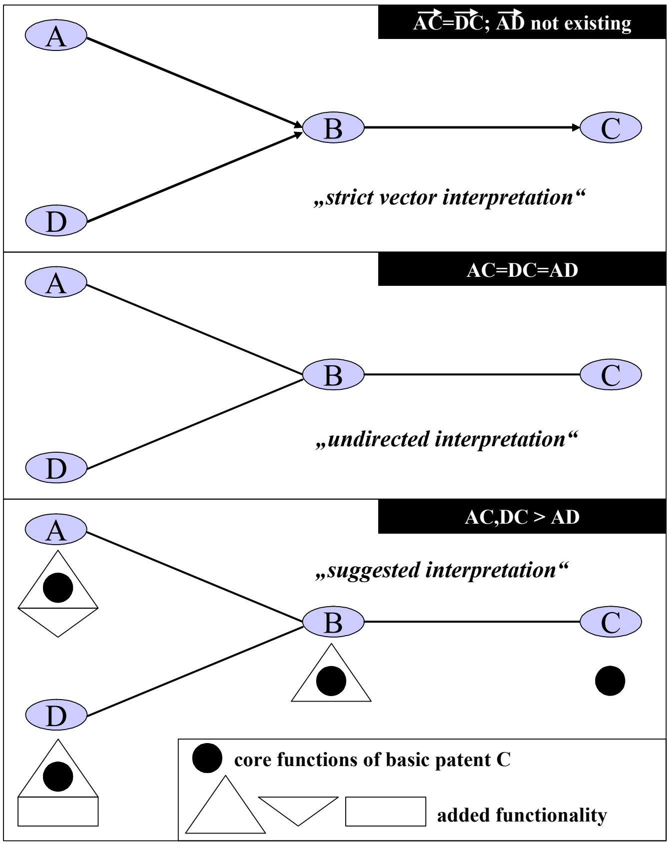



**Figure 4: Interpretation of indirect citations and bibliographical coupling**

A differentiated perspective is needed, if the direction changes between single stages of the citation flow (Figure 4). Given that patents A, D cite B, and B cites C, we retrieve the unidirectional relationships A->B->C as well as D->B->C. A strict vector interpretation on relationship patterns would suggest the absence of any relationship between A and D. Otherwise, neglecting the vector direction, one would conclude about identical proximities of A-C (given by A-B-C) and A-D (A-B-D). Both interpretations are inadequate: On the one hand, A and D are somehow related as they both cite B. On the other hand, it is reasonable to assume that A,D encompass basic functions of C, whereas the specific functions of A and D, which go above and beyond of those documented in C should be unrelated because we do not observe any citation linkage between these two more recent patents. Thus, the link A-D should be of lesser value than A-C or D-C, but not zero. In our analysis, we value such an indirect link between two patents which entails a biographical coupling as equal to ½ of the value of an one-way link. For simplification, we neglect any relationship between two patents containing more than one stage of biographical coupling. This seems reasonable, since such complex relationship patterns are hard to assess and may well be misleading, as they may be based on specific technological details which are only loosely related to the core of the patented invention.

# 3 Application to the field of mobile telecommunication (the H04Q-007 Blob)

## 3.1 The Task: The H04Q-007-Technology Blob

The main difficulties using the IPC for patent analysis were mentioned above. However, the true "pièce de resistance" for both classification schemes is the fact that patent applications and issued patents in the field of mobile telecommunication are grouped into only one main group: H04Q-007.

Mobile telecommunication is almost a perfect example to show the weaknesses and limitations of the IPC: While the basic radio communication technology did not change significantly, the treatment of the transmitted data was completely altered. The first major shift occurred when voice data was transmitted digitally. This opened the door to a host of additional services, ranging from the now famous SMS to other data-centered applications and integration with consumer electronics. Apart from technical advancement in the communication integrated circuits themselves, innovation now focuses more on these additional layers - from software, network management to providers and integration.

Furthermore, the relatively small number of subgroups is not equivalent to apparent products and applications. This is confirmed by looking at Table 1, which contains the relevant subgroups of the IPC for mobile telecommunications.

**Table 1: Categories within H04Q-007**

| |
|---|
| 7/00 **Selecting arrangements to which subscribers are connected via radio links or inductive links** |
| 7/20 . in which the radio or inductive links are two-way links, e.g. mobile radio systems [6] |
| 7/22 . . using dedicated mobile switching centres, e.g. cellular systems [6] |
| 7/24 . . using public exchanges or networks with at least partially integrated mobile switching or mobile application [6] |
| 7/26 . . using a private branch exchange (PBX) as final selecting device, e.g. cordless PBX [6] |
| 7/28 . . Trunked radio systems, i.e. sharing radio channel among active subscribers [6] |
| 7/30 . . Base station equipment [6] |
| 7/32 . . Mobile subscriber equipment [6] |
| 7/34 . . Test or monitoring equipment [6] |
| 7/36 . . Arrangements for mobile service area coverage, e.g. cells layout [6] |
| 7/38 . . Arrangements for completing call to or from mobile subscriber [6] |

Although being more up to date and more detailed especially in the field of mobile telecommunications, the European classification (ECLA) follows the same spirit as the IPC. Indeed, we are confronted with one large "technology blob" when we look closely at this category[28]. The blob contains approximately 90.000 patent documents[29]. The existence of that technology blob impedes the usage of either one of the three methods for categorizing patents:

- Relying on predefined IPC/EC-main groups is out of question because there is just one IPC main group (H04Q-007) that contains the vast majority of patents relating to the mobile telecommunication industry. Furthermore, the relatively small number of subgroups is not equivalent to apparent products and applications.

- Manual categorization based on textual information in patent abstracts as well as in full text archives by using Boolean query operators are put to question, given that the external researcher does not have an ex-ante semantic list of concepts he should mine for.

[28] A "blob" is a lump of something ill-defined or amorphous.

[29] Source: Espacenet worldwide collection.

- Recently, research has shown that automated text classification can produce meaningful results (Teichert and Mittermayer 2002, Brücher et al. 2002). No matter how sophisticatedly one applies intelligent text mining tools, a basic tenet remains: Text mining always looks at one document (patent) at a time. Text clustering tools try to group the documents according to their similarity. Similarity is what they were trained to look for by using text pre-processing procedures. Thus, information about patterns of time effects cannot be brought out of automated text mining algorithms since they perform time independent 'textual similarity hunting'.

What is required is a different approach to technology classification, one that groups patents into categories that are more accessible to economically relevant analysis. Such a method may be based on 'reference hunting' (Bendl and Weber 2002) because therein one might reveal evolutionary (path dependent) patterns of technical change along technical trajectories (Dosi 1982) and technology cycles (Tushman and Anderson 1986, Tushman et al. 1997). In the following, we test this alternative method for "slicing the blob" using citation information.

## 3.2 Implementation

### 3.2.1 Database

We link two data repositories of the European Patent Office (EPO): The European Patent Register (EUREG) and the EPODOC databases[30]. The citations used in our analysis are found in the search reports of the corresponding patent or patent applications.

For this reason, we entered a partnership with the Swiss Federal Institute for Intellectual Property. All European patent applications filed with the EPO by applicants domiciled in Europe classified in the IPC main group H04Q-007 were taken into account.[31] From the 6309 patent documents in the mobile telecommunications sector, the vast majority contains citation information. We further group these patent documents into patent families which relate to one

[30] While the file EUREG contains the legal status of patent applications filed with the European patent office, EPODOC is focused on its technical content, and is used by European examiners for their prior art searches.

[31] According to Michel and Bettels "for the researcher envisaging comparative studies, comparison of citations obtained within one system […] is therefore undoubtedly more significant than comparison of data from legally different patenting systems." (Michel and Bettels 2001, p. 194).

and the same invention.[32] The citation connections between them are making up the adjacency matrix which shows the citation connections between patent families.[33]

### 3.2.2 Generated algorithms

The calculation of the technological connectivity of the patents occurred along the description specified in chapter 2.3. Figure 5 displays the conception: core patents of the H04Q007 range were differentiated from collective third party patents. Collective third party patents result from the bibliographic coupling of the patents occurring outside of the IPC class H04Q007.

The determination of the technological connectivity of the patents took place in a multistage procedure divided into two parts. Figure 6 contains the constructive steps of the analysis of the example of Figure 5 in a simplified graphical and mathematical form. The directional relations of the patents among each other are evaluated in one strand (A, B. E, H), in the other strand (A, C, F, G) the bibliographic couplings are examined. In a first step, the incidence matrix (A) of all patents is entered, whereby the core patents form the rows of the matrix and the cited patents form the columns. The sum of the rows equals the number of cited patents per patent; this result should then be added as an indicator of the value as mentioned earlier. To enable the usage of this information on multiplicative models an attribute-matrix (D) with the sum of the quotations as the diagonal vector was formed. Furthermore, the incidence matrix was split into a socio-matrix (B) and an affiliation-matrix (C). The socio-matrix (B) contains the number of quotations of the core patents among each other; the affiliation-matrix (C) contains the number of “collective events” or also the number of collective quotations of third-party patents (bibliographic coupling). Both matrices are subsequently multiplied with the attribute-matrix (D) to help define the technological proximity of two patents. Technical reasons then make the multiplication of the affiliation-matrix (F) with its transpose (F’) necessary thereby retaining the symmetric probabilities of accessibility of two patents, and also the multiplication with the value 2 (G), allowing the consideration of the smaller proximity of the bibliographic coupling.

---

[32] In technical terms, patent families are defined as a group of patents with the same or related priority information.

[33] “Network analysts commonly think of their data as *graphs*. A graph is a set of points (also known as nodes or vertices) together with a set of lines (links, ties, edges) that connect the points. The information in a graph (who is connected to whom) can be represented by a matrix known as the *adjacency matrix*, in which a given cell X(i,j) contains a value of 1 if nodes *i* and *j* are connected, and 0 otherwise.” (Borgatti 2002, User Manual, p. 15)

In the case of the socio-matrix (B), the indirect connections of the patents should also be integrated. Out of these reasons, the *reachability* of the probabilities-matrix (H) shall be determined. This step depends on the determination of the most probable or also of the optimal path between two patents. In this connection, the matrix is calculated to the power of itself until the probability of all the elements is 0. The result is the asymmetric matrix (H), which is then made maximally symmetrical and afterwards added to the matrix G. The obtained influence-matrix (I) summarizes the results of the first strand (A, B. E, H), which contains the vectored relations of the patents among each other, as well as the results of the second strand with the bibliographic couplings (A, C, F, G).

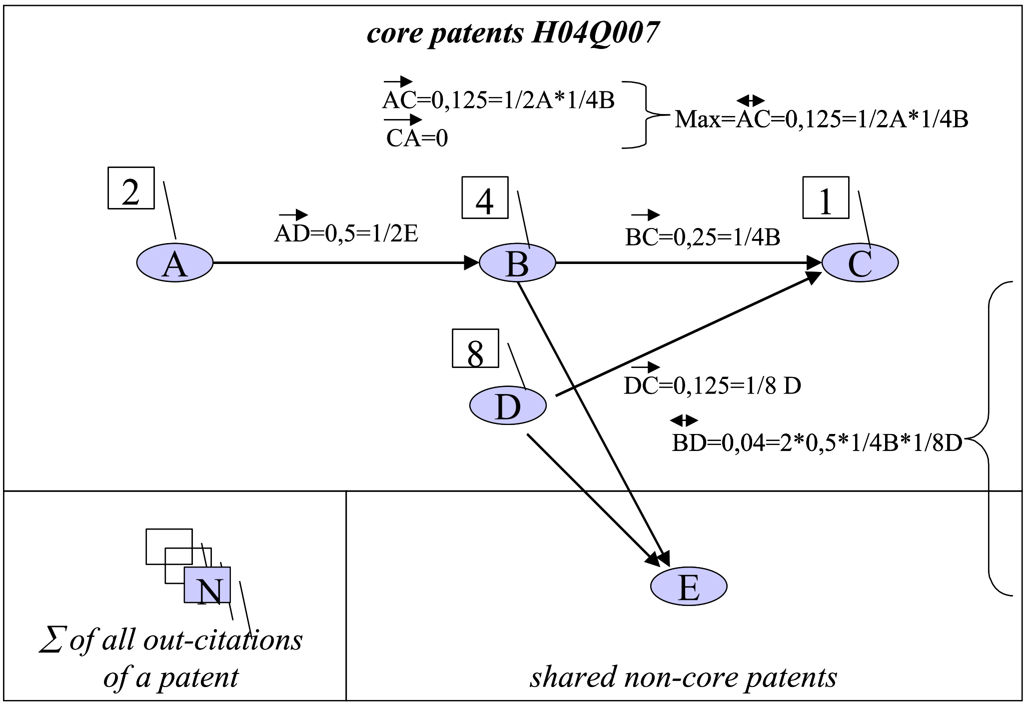



**Figure 5: Example for Vector Information**

***A: Incidence Matrix (two-mode network)***
*$I_{ij}=iRj=g*h$ mit $h_{ij}=(n_i,m_j)$*

| | A | B | C | D | E | ... | Σ |
|---|---|---|---|---|---|---|---|
| **A** | - | 1 | 0 | 0 | 0 | | 2 |
| **B** | 0 | - | 1 | 0 | 1 | | 4 |
| **C** | 0 | 0 | - | 0 | 0 | | 1 |
| **D** | 0 | 0 | 1 | - | 1 | | 8 |

***B: Sociomatrix***
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | - | 1 | 0 | 0 |
| **B** | 0 | - | 1 | 0 |
| **C** | 0 | 0 | - | 0 |
| **D** | 0 | 0 | 1 | - |

***E: Probabilities***
*E = D*B; i=j=1*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | 1 | 1/2 | 0 | 0 |
| **B** | 0 | 1 | 1/4 | 0 |
| **C** | 0 | 0 | 1 | 0 |
| **D** | 0 | 0 | 1/8 | 1 |

***H: Reachability***
*$X^p=\{x_{ij}^{\{\Sigma p\}}\}$ mit $\{X^p, 1\leq p \leq g-1\}$*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | 1 | 1/2 | 1/8 | 0 |
| **B** | 0 | 1 | 1/4 | 0 |
| **C** | 0 | 0 | 1 | 0 |
| **D** | 0 | 0 | 1/8 | 1 |

***C: Affiliations Matrix***
*$X_{ij}=I_{ij*}I'_{ij}$ mit $i'_{ij}=i_{ji}$*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}=x_{ji}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | - | 0 | 0 | 0 |
| **B** | 0 | - | 0 | 2 |
| **C** | 0 | 0 | - | 0 |
| **D** | 0 | 2 | 0 | - |

***D: Attribut Matrix***
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}=x_{JI}$; i=j=1/n*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | 1/2 | - | - | - |
| **B** | - | 1/4 | - | - |
| **C** | - | - | 1/1 | - |
| **D** | - | - | - | 1/8 |

***I: Influence Matrix***
*Σ(H,G) mit $x_{ij}=x_{JI}$=max*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}=x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | 1 | 1/2 | 1/8 | 0 |
| **B** | 1/2 | 1 | 1/4 | 4/64 |
| **C** | 1/8 | 1/4 | 1 | 1/8 |
| **D** | 0 | 4/64 | 1/8 | 1 |

***F: Probabilities***
*F = D*C*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | - | 0 | 0 | 0 |
| **B** | 0 | - | 0 | 2/4 |
| **C** | 0 | 0 | - | 0 |
| **D** | 0 | 2/8 | 0 | - |

***F': Transpose of F***
*F' mit $x'_{ij}=x_{ji}$*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | - | 0 | 0 | 0 |
| **B** | 0 | - | 0 | 2/8 |
| **C** | 0 | 0 | - | 0 |
| **D** | 0 | 2/4 | 0 | - |

***G: Reachability***
*G=(F* F')*1/2; ; i=j=0*
*$X_{ij}=iRi=g*g$ mit $x_{ij}=(n_i,n_j)$; $x_{ij}=x_{JI}$*

| | A | B | C | D |
|---|---|---|---|---|
| **A** | 0 | 0 | 0 | 0 |
| **B** | 0 | 0 | 0 | 4/64 |
| **C** | 0 | 0 | 0 | 0 |
| **D** | 0 | 4/64 | 0 | 0 |

***Source:*** *$I_{ij}=iRj=g*h$ with $h_{ij}=(n_i,m_j)$ ->two-mode sociomatrix;*
*$X_{ij}=iRi=g*g$ with $x_{ij}=(n_i,n_j)$ -> one-mode sociomatrix*
*$X_{ij}=iRi=g*g$ with $x_{ij}=(n_i,n_j)$; $x_{ij}=x_{JI}$ -> symmetric one-mode sociomatrix;*
*$X_{ij}=iRi=g*g$ with $x_{ij}=(n_i,n_j)$; $x_{ij}\neq x_{JI}$ -> not symmetric one-mode sociomatrix*
*n = Σ of all out-citations of a patent*

**Figure 6: Generated Algorithms**

### 3.2.3 Grouping

The matrix I represents the input for the procedure of the hierarchical cluster analysis. The influence-matrix was resolved of unlinked patents before a determination of the clusters took place. The input matrix for the cluster analysis contained 3131 of originally 4279 patent families, which amounts to about 27% of unlinked patents in the IPC-class H04Q007. An examination of the age of these patents clearly indicated a majority of young patents, on which a inspection report does not yet exist and therefore citations are not yet possible.

The implemented clustering procedure is based on a hierarchical procedure and the method of average similarity. The selection of the optimal number of clusters occurred over the criteria of group sizes and stability. 14 stable and from the size of the group meaningful clusters were identified. These clusters contain 2341 patents; 75% of the cases in this procedure were able to be grouped. The following evaluation is based on the determined groups, whereby the clusters are interpreted as technological fields.

The quality of our determined groups is analyzed with their comparison to the IPC-classification. As illustrated in Table 2, the IPC-classification is only capable of a discrimination of maximally 3 groups. The classes H04Q7/38, H04Q7/22 und H04Q7/32 alone contain nearly 60% of all the cases (see Table 2). This is definitely less differentiated than our groups and does not comply with the goal of a technical classification of the mobile telephone sector that is refined enough. As can further be seen in Table 2, the IPC-classes have only a limited explanatory content for the obtained clusters. And so the largest 6 listed IPC-classes have too many overlaps to be able to explain the differences of the groups. This underlines the necessity of out following analysis.

**Table 2: Technological Clusters and IPC-Classification**

| ***Cluster/ IPC*** | *H04Q7/38* | *H04Q7/22* | *H04Q7/32* | *H04Q7/24* | *H04Q7/36* | *H04Q7/30* | *H04Q7/34* | ***sum*** | ***Overall N (100%)*** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 16,7% | 14,0% | 31,2% | 2,9% | 0,7% | 0,2% | 1,6% | 67,3% | 449 |
| 2 | 25,5% | 32,2% | 4,6% | 3,9% | 0,8% | 2,1% | 0,0% | 69,1% | 388 |
| 3 | 36,6% | 25,2% | 6,7% | 2,4% | 1,2% | 5,9% | 1,2% | 79,1% | 254 |
| 4 | 51,2% | 14,3% | 5,2% | 1,2% | 2,8% | 0,8% | 0,4% | 75,8% | 252 |
| 5 | 62,4% | 8,5% | 2,8% | 2,8% | 1,4% | 0,7% | 0,0% | 78,7% | 141 |
| 6 | 23,1% | 0,8% | 3,1% | 0,0% | 11,5% | 3,1% | 0,8% | 42,3% | 130 |
| 7 | 43,5% | 13,9% | 5,2% | 4,3% | 0,9% | 5,2% | 0,0% | 73,0% | 115 |
| 8 | 56,9% | 6,9% | 2,9% | 0,0% | 2,0% | 1,0% | 2,9% | 72,5% | 102 |
| 9 | 40,6% | 5,9% | 3,0% | 3,0% | 9,9% | 5,0% | 9,9% | 77,2% | 101 |
| 10 | 18,0% | 4,0% | 15,0% | 1,0% | 0,0% | 5,0% | 2,0% | 45,0% | 100 |
| 11 | 41,1% | 3,3% | 18,9% | 1,1% | 0,0% | 0,0% | 0,0% | 64,4% | 90 |
| 12 | 36,7% | 31,1% | 5,6% | 0,0% | 0,0% | 0,0% | 2,2% | 75,6% | 90 |
| 13 | 38,0% | 2,5% | 8,9% | 0,0% | 12,7% | 0,0% | 5,1% | 67,1% | 79 |
| 14 | 24,0% | 8,0% | 0,0% | 14,0% | 0,0% | 10,0% | 0,0% | 56,0% | 50 |
| **Sum** | ***33,9%*** | ***15,8%*** | ***10,8%*** | ***2,5%*** | ***2,4%*** | ***2,3%*** | ***1,4%*** | ***69,0%*** | **2341** |

### 3.2.4 Interpretation of the groups

A denomination of these technological fields and a first validation of the findings is made possible by a statistical content analysis. For this, all the words are selected from the English patent titles. After an adjustment of expletives (the, and, or…) the 300 words occurring most often are divided into variables and their percentage of appearance within the clusters is evaluated. To enable inter-comparability, the number of words is standardized by the corresponding size of the group. In the following, these obtained variables are integrated into a principal component analysis. The requirements of the model allow only 13 factors. Table 3 shows that the clusters load without overlaps onto the determined dimensions. This puts us into the position to be able to identify words with significantly different frequencies of occurrence in the patent titles. A preliminary, rough and contextual validation of the obtained groups is postulated to this end.

**Table 3: Principal Components Analysis of Patent Titles and Derived Technological Clusters**

| **Factor** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Eigenvalue (unrotated)* | *49,1* | *37,7* | *30,7* | *25,4* | *22,7* | *20,5* | *17,9* | *15,1* | *13,1* | *10,5* | *9,0* | *7,0* | *2,3* |
| *Eigenvalue (rotated)* | *44,5* | *30,6* | *28,7* | *23,8* | *21,1* | *18,1* | *17,5* | *14,5* | *13,6* | *12,4* | *10,5* | *8,9* | *3,6* |
| ***Cluster/ Factor loadings*** | | | | | | | | | | | | | |
| *1* | -0,31 | -0,21 | -0,32 | -0,36 | -0,30 | 0,34 | -0,44 | -0,25 | -0,22 | 0,20 | **3,34** | 0,04 | -0,06 |
| *2* | -0,45 | -0,61 | -0,38 | -0,54 | -0,58 | 0,72 | -1,05 | -0,64 | -0,69 | 0,91 | -0,81 | -1,46 | **2,14** |
| *3* | -0,34 | -0,70 | -0,29 | -0,20 | -0,28 | 0,52 | -0,48 | -0,52 | -0,61 | 0,51 | -0,69 | **3,06** | -0,04 |
| *4* | -0,40 | 0,97 | -0,36 | -0,47 | -0,54 | 0,43 | -0,63 | -0,44 | -0,76 | 0,79 | -0,68 | -0,96 | **-2,64** |
| *5* | -0,03 | **3,17** | -0,22 | -0,03 | -0,13 | -0,19 | 0,50 | 0,05 | 0,04 | -0,16 | -0,01 | 0,55 | 1,16 |
| *6* | -0,19 | -0,17 | -0,18 | **3,44** | -0,14 | 0,23 | -0,15 | -0,06 | -0,03 | -0,02 | -0,01 | -0,24 | -0,06 |
| *7* | -0,14 | -0,34 | -0,25 | -0,28 | -0,26 | 0,21 | -0,08 | **3,40** | 0,02 | 0,25 | -0,18 | -0,01 | -0,01 |
| *8* | -0,26 | -0,46 | -0,21 | -0,12 | -0,15 | **-3,40** | -0,22 | -0,18 | -0,17 | 0,24 | -0,10 | -0,07 | 0,00 |
| *9* | -0,38 | -0,73 | -0,30 | -0,26 | -0,37 | 0,30 | **3,25** | -0,40 | -0,29 | 0,27 | -0,11 | -0,33 | -0,04 |
| *10* | -0,09 | -0,03 | **3,47** | -0,11 | -0,08 | 0,09 | -0,03 | -0,05 | -0,01 | 0,07 | 0,00 | -0,05 | -0,01 |
| *11* | -0,26 | -0,22 | -0,30 | -0,34 | -0,24 | 0,26 | -0,26 | -0,44 | **3,35** | 0,13 | -0,28 | -0,04 | -0,21 |
| *12* | -0,33 | -0,35 | -0,25 | -0,41 | -0,32 | 0,22 | -0,33 | -0,19 | -0,39 | **-3,32** | -0,34 | -0,22 | -0,12 |
| *13* | -0,27 | -0,09 | -0,21 | -0,19 | **3,43** | 0,21 | -0,05 | -0,09 | -0,14 | 0,09 | -0,09 | -0,20 | -0,05 |
| *14* | **3,45** | -0,23 | -0,19 | -0,12 | -0,05 | 0,08 | -0,02 | -0,19 | -0,09 | 0,03 | -0,04 | -0,09 | -0,07 |

## 3.3 Discussion

A provisional naming of the clusters occurs over the words corresponding most to the factor (specific factor load). The applied nomenclature is then specified and verified using the abstract and the available literature concerning the technologies of the mobile telephoning sector as a base. Figure 7 visualizes the results. As discernable, a rough differentiation of cellular networks of long-range distance and mobile telephoning and cellular networks of the short-range distance is possible. Cellular networks of the long-range distance contain for example walksystems, Wireless installations, paging networks, trunked mobile radio systems, advanced radio data information services and localization techniques along with 2-G and 3-G technologies. In addition to this, a technological area was determined that contains base techniques for a wireless transmission, such as multiplex procedures and frequency modulation, network connection. A final large sector is formed by technologies for mobile information systems, such as software, basic functions of application devices or protocols, data transmission etc.

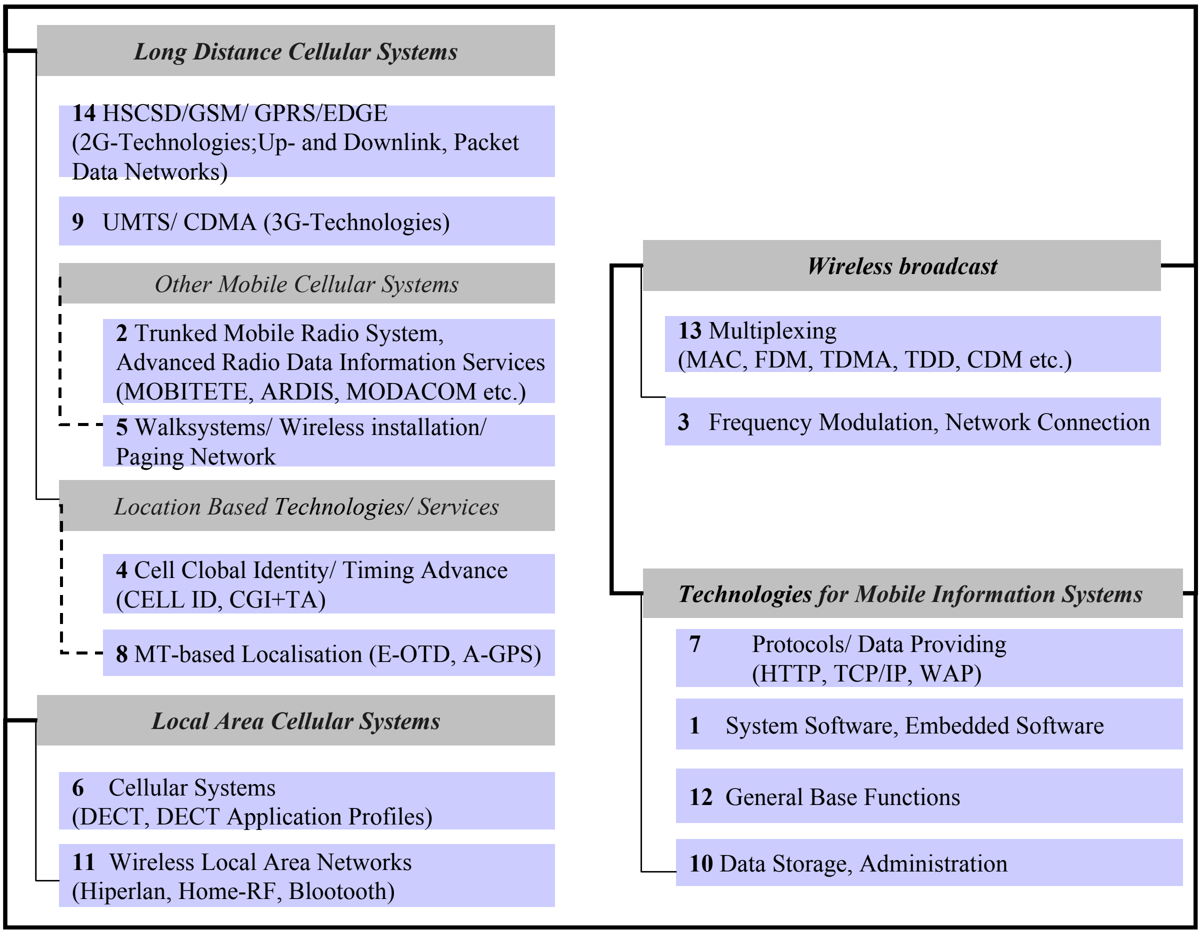



**Figure 7: Interpretation of Derived Technological Clusters**

If the identified groups are examined concerning to technological interweavements, they prove to have very few overlapping elements (see Table 4). The strongest reflexive reference (1,47) is observable in the wireless local networks. The whole area of short-range cellular networks is characterized furthermore by a relatively strong technological interweavement with other core sectors of the mobile telephone market. This result appears plausible as the solutions for the short-range sector require a strong technological integration.

**Table 4: Dependency Patterns of derived Technological Clusters**

| *Cluster Density ≥ 0,02* | | *Long Distance* | | | | | | *Local Area* | | *Broadcast* | | *Information Technologies* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 14 | 9 | 2 | 5 | 4 | 8 | 6 | 11 | 13 | 3 | 7 | 1 | 12 | 10 |
| ***Long Distance*** | 14 | 0,20 | | | | | 0,02 | | 0,02 | | | | | | |
| | 9 | | 0,27 | | 0,02 | | | 0,02 | 0,02 | | | 0,03 | | | |
| | 2 | | | 0,24 | | 0,02 | | 0,02 | 0,03 | | | | | | 0,02 |
| | 5 | | 0,02 | | 0,60 | | | 0,04 | | | | | | | |
| | 4 | | | 0,02 | | 0,25 | 0,03 | 0,03 | | | 0,02 | | | | |
| | 8 | 0,02 | | | | 0,03 | 0,28 | | 0,04 | | 0,02 | 0,03 | 0,02 | | |
| ***Local Area*** | 6 | | 0,02 | 0,02 | 0,04 | 0,03 | | 0,29 | 0,02 | 0,03 | | | 0,03 | 0,02 | |
| | 11 | 0,02 | 0,02 | 0,03 | | | 0,04 | 0,02 | 1,47 | 0,02 | 0,03 | 0,04 | 0,03 | 0,03 | |
| ***Broadcast*** | 13 | | | | | | | 0,03 | 0,02 | 0,35 | 0,03 | | | | 0,02 |
| | 3 | | | | | 0,02 | 0,02 | | 0,03 | 0,03 | 0,25 | 0,02 | | | |
| ***Information*** | 7 | | 0,03 | | | | 0,03 | | 0,04 | | 0,02 | 0,31 | 0,03 | | |
| ***Technologies*** | 1 | | | | | | 0,02 | 0,03 | 0,03 | | | 0,03 | 0,27 | | 0,02 |
| | 12 | | | | | | | 0,02 | 0,03 | | | | | 0,22 | |
| | 10 | | | 0,02 | | | | | | 0,02 | | | 0,02 | | 0,36 |

Table 5 contains detailed information of the obtained technological clusters. Indications of the number of patents within the group, of the average age of the patents are supplied along with the average number of citations per patent (an adjustment by the age of the patent has been omitted) and the obtained value of dependence of the influence-matrix (I) has been displayed. As becomes apparent, the mean value of the citations of the patents of one technological field is not in context with its technological entwinement, which considers the total of all direct and indirect relationships. Therefore the frequency of quotation only possesses a limited explanatory power, even though it is often used as an indicator for the quality of information. Cluster 3 (frequency modulation, network connection) for example differs from cluster 8 (MT based localizations techniques) not in regard to the average number of quotations per patent but the former cluster contains patents which are more strongly integrated into the complete mobile network.

**Table 5: Measurements of derived Technological Clusters**

| *Cluster.* | *Technological Field* | *number of patents* | *average age* | *average citation/ patent* | *average degree/ patent* | *Percent of Big Market Player* | *Percent of Small Market Player* |
|---|---|---|---|---|---|---|---|
| 1 | *System Software, Embedded Software* | 449 | 1998,6 | 1,54 | 23,0 | 78,9% | 21,1% |
| 2 | *Trunked Mobile Radio System, Advanced Radio Data Information Services (MOBITETE, ARDIS, MODACOM etc.)* | 388 | 1999,3 | 1,34 | 14,9 | 88,2% | 11,8% |
| 3 | *Frequency Modulation, Network Connection* | 254 | 1998,2 | 1,52 | 78,4 | 89,9% | 10,1% |
| 4 | *Cell Clobal Identity/ Timing Advance (CELL ID, CGI+TA)* | 252 | 1998,0 | 1,58 | 15,2 | 93,3% | 6,7% |
| 5 | *Walksystems/ Wireless Installation/ Paging Network* | 141 | 1998,7 | 1,57 | 15,2 | 95,5% | 4,5% |
| 6 | *Cellular Systems (DECT, DECT Application Profiles)* | 130 | 1993,1 | 1,55 | 8,9 | 85,5% | 14,5% |
| 7 | *Protocols/ Data Providing (HTTP, TCP/IP, WAP)* | 115 | 1998,8 | 1,31 | 10,5 | 92,5% | 7,5% |
| 8 | *MT-based Localisation (E-OTD, A-GPS)* | 102 | 1997,7 | 1,58 | 11,0 | 80,7% | 19,3% |
| 9 | *UMTS/ CDMA (3G-Technologies)* | 101 | 1998,1 | 1,36 | 6,7 | 94,3% | 5,7% |
| 10 | *Data Storage, Administration* | 100 | 1996,4 | 1,45 | 8,7 | 83,9% | 16,1% |
| 11 | *Wireless Local Area Networks (Hiperlan, Home-RF, Bluetooth)* | 90 | 1997,3 | 1,29 | 10,1 | 91,4% | 8,6% |
| 12 | *General Base Functions* | 90 | 1999,0 | 1,19 | 5,7 | 80,3% | 19,7% |
| 13 | *Multiplexing (MAC, FDM, TDMA, TDD, CDM etc.)* | 79 | 1997,7 | 1,10 | 5,4 | 91,8% | 8,2% |
| 14 | *HSCSD/GSM/ GPRS/EDGE (2G-Technologies;Up- and Downlink, Packet Data Networks)* | 50 | 1997,2 | 1,40 | 9,6 | 90,2% | 9,8% |

Finally, Table 5 portrays indicatively the rate of patenting of large versus small market participators within the obtained technology fields. The distinction took place via the notoriety of the applicant, the size of the firm and the relative technology focus on the mobile telephone sector. Larger market participants dominate according to this clearly in the sector of 2-G and 3-G technologies and wireless technologies. This leadership in technology in the long-range sector seems inevitable in the perspective of the fact that a market entry is practically futile for small providers.

# 4 Summary

Studies of technological change constitute a research field of growing importance and sophistication. State-of-the-art methods of patent analysis go well beyond simple patent counts and utilize patent citation analysis as a means to reveal information on the complex, interrelated and cumulative processes of technological change. This allows assessments not only on single patented inventions but as well on the long-term innovative positioning of entire entities, as companies, clusters or nations.

We strive to contribute to the current discussion with an in-depth methodological reflection on the potential of patent citation network analysis. Our analyses base on an in-depth evaluation of the information sources used. Pointing to the very specific patterns of patent citation information, we conclude that single-stage citation analyses are insufficient for revealing specific paths of technological development. To mirror actual developments in a certain technological field, citation analysis should rely on all, bibliographical coupling, co-citations, direct and indirect citations.

Furthermore, we explain that standard measures of network analysis, as used for describing social entities, have to be adjusted to the specific conditions and purposes of patent analysis. As a consequence of our methodological considerations, we introduce a novel approach for assessing technological proximity based on adjusted measures of reachability. We specify rules for measuring both direct and indirect citation links and propose a way to deal with directional changes as manifested in bibliographical couplings.

To exemplify the feasibility of our approach, a large-scale application to the field of mobile telecommunication (the H04Q-007 Blob) is provided. Using hierarchical cluster analysis based on citation information that is collected within a relational database we can ‘capitalize’ on all of the citation types mentioned above. Solely by means of patent citation network analysis, we are able to identify fourteen technology fields using a hierarchical cluster analysis.

To cross-check the meaningfulness of our derived classification, we apply different methods which help to assess the validity of the achieved findings. The analyses univocally enforce our findings. Firstly, the statistical clustering represents the different layers of the mobile telecommunications industry, and is thus of high face validity. Secondly, the revealed technology clusters are consistent with the findings of a separate principal component analysis based on textual information (key words), which indicates an external validity. Finally, a draft context analysis reveals an intuitively appealing competitive profiling of the derived technology fields. Thus, we conclude about the high potential benefits of patent citation network analysis and encourage for further research in this area.

# 5 References

Akers, N.J., 2000. The Referencing of Prior Art Documents in European Patents and Applications. World Patent Information 22, 309-315.

Albert, M.B., Avery, D., Narin, F., McAllister P., 1991. Direct Validation of Citation Counts as Indicators of Industrially Important Patents. Research Policy 20, 251-259.

Almeida, P., 1996. Knowledge Sourcing by Foreign Multinationals: Patent Citation Analysis in the US Semiconductor Industry. Strategic Management Journal 17, 155–165.

Archibugi, D., 1992. Patenting as an Indicator of Technological Innovation: A Review. Science and Public Policy 19, 357–368.

Arthur, W.B., 1989. Competing Technologies, Increasing Returns, and Lock-In by Historical Events. The Economic Journal 99, 116-131.

Arundel, A., Kabla, I., 1998. What Percentage of Innovations are Patented? Empirical Estimates for European Firms. Research Policy 27, 127-141.

Basberg, B., 1987. Patents and Measurement of Technological Change: A Survey of the Literature. Research Policy 16, 131–141.

Bendl, E., Weber, G., 2002. Patenrecherche und Internet, Heymanns, Köln.

Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard: Analytic Technologies.

Breitzman, A., Thomas, P., 2002. Using Patent Citation Analysis to Target/Value M&A Candidates. Research Technology Management 45, 28-46.

Brücher, H., Knolmayer, G., Mittermayer, M-A., 2002. Document Classification Methods for Organizing Explicit Knowledge. Proceedings of the Third European Conference on Organizational Knowledge, Learning and Capabilities, Athens 2002.

Carpenter, M.P., Narin, F., Woolf, P., 1981. Citation Rates to Technologically Important Patents. World Patent Information 3, 160-163.

Cohen, W.M., Levinthal, D.A., 1989. Innovation and Learning: The two Faces of R&D. The Economic Journal 99, 569-596.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly 35, 128-152.

Cohen, W.M., Levinthal, D.A., 1994. Fortune Favors the Prepared Firm. Management Science 40, 227-251.

Debackere, K., Verbeek, A., Luwel, M., Zimmerman, E., 2002. Measuring Progress and Evolution in Science and Technology II: The Multiple Uses of Technometric Indicators. International Journal of Management Reviews 4, 213-231.

Deng, Z., Lev, B., Narin, F., 1999. Science and Technology as Predictors of Stock Market Performance. Financial Analysts Journal 55, 20-32.

Dosi, G., 1982. Technological Paradigms and Technological Trajectories. Research Policy 11, 147-162.

Espacenet Worldwide Collection: see, e.g. http://ch.espacenet.com/espacenet/ch/en/e_net.htm, and enter H04Q7 in the IPC classification field.

Fleming, L., Sorenson, O., 2001. Technology as a Complex Adaptive System: Evidence from Patent Data. Research Policy 30, 1019–1039.

Griliches, Z., 1990. Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature 28, 1661–1707.

Grupp, H., 1998. Foundations of the Economics of Innovation. Edward Elgar, Cheltenham/Northhampton, MA.

Guidelines for the Examination in the EPO: http://www.european-patent-office.org/legal/gui_lines/e/index.htm

Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498. National Bureau of Economic Research, Cambridge, MA.

Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation Frequency and the Value of Patented Inventions, Review of Economics & Statistics 81, 511-515.

Harhoff, D., Reitzig, M., 2001. Strategien zur Gewinnmaximierung bei der Anmeldung von Patenten, Wirtschaftliche und rechtliche Aspekte als Entscheidungsgrößen beim Schutz von FuE. Zeitschrift für Betriebswirtschaft 71, 509-530.

Harhoff, D., Scherer, F.M., Vopel, K., 2002. Citations, Family Size, Opposition and the Value of Patent Rights, Research Policy, In Press, Corrected Proof, Available online 4 December 2002, 1-21.

Henderson, R., Jaffe, A., Trajtenberg, M., 1998. Universities as a Source of CommercialTechnology: A Detailed Analysis of University Patenting 1965-1988. Review of Economics and Statistics 80, 119-127.

Jaffe, A. B., 1998. Patents, Patent Citations, and the Dynamics of Technological Change. NBER Reporter (Summer), 8-11.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. Quarterly Journal of Economics 108, 577-598.

Lanjouw, J.O., Pakes, A., Putnam, J., 1998. How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data. Journal of Industrial Economics XLVI (4), 405–433.

Lerner, J., 1994. The Importance of Patent Scope: An Empirical Analysis. RAND Journal of Economics 25, 319-333.

Mansfield, E., 1986. Patents and Innovation: An Empirical Study. Management Science 32, 173–181.

Michel, J., Bettels, B., 2001. Patent Citation Analysis, A Closer Look at the Basic Input Data form Patent Search Reports. Scientometrics 51, 185-201.

Mogee, M.E., Kolar, R.G., 1994. International Patent Analysis as a Tool for Corporate Technology Analysis and Planning. Technology Analysis & Strategic Management 6, 485-503.

Narin F., Noma, E., Perry, R., 1987. Patents as Indicators of Corporate Technological Strength. Research Policy 16, 143-155.

OECD, 1994. The Measurement of Scientific and Technological Activities. Using Patent Data as Science and Technology Indicators, Patent Manual. OECD, Paris.

Pakes, A., 1986. Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. Econometrica 54, 755-784.

Pakes, A., Schankerman, M., 1984. The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources. In: Griliches, Z. (Ed.), R&D, Patents, and Productivity. University of Chicago Press, Chicago, 73–88.

Pavitt, K., 1985. Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems. Scientometrics 7, 77–79.

Pilkington, A., Dyerson, R., Tissier, O., 2002. The Electric Vehicle: Patent Data as Indicators of Technological Development. World Patent Information 24, 5-12.

Rosenberg, N., 1976. Perspectives on Technology, Cambridge University Press, Cambridge.

Rosenberg, N., 1982. Inside the Black Box: Technology and Economics. Cambridge University Press, Cambridge.

Rosenkopf, L., Nerkar, A., 2001. Beyond Local Search: Boundary-spanning, Exploration, and Impact in the Optical Disc Industry. Strategic Management Journal 22, 287-306.

Scherer, F.M., 1982. Inter-Industry Technology Flows and Productivity Growth. Review of Economics and Statistics 64, 627-634.

Scherer, F.M., 1983. The Propensity to Patent. International Journal of Industrial Organization 1, 107–128.

Schmokler, J., 1966. Invention and Economic Growth. Harvard University Press, Cambridge.

Schwander, P., 2000. An Evaluation of Patent Searching Resources: Comparing the Professional and Free-Online Databases. World Patent Information 22, 147-165.

Soete, L.L.G., Wyatt, S.M.E., 1983, The Use of Foreign Patenting as an Internationally Comparable Science and Technology Output Indicator, Scientometrics 5, 31-54.

Stuart, T.E., Podolny, J.M., 1996. Local Search and the Evolution of Technological Capabilities. Evolutionary Perspectives on Strategy. Strategic Management Journal 17, 21-38.

Teece, D.J., 1986. Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy. Research Policy 15, 285-305.

Teichert, Th., Mittermayer, M-A., 2002. Text Mining for Technology Monitoring. Proceedings of the 2002 IEEE International Engineering Management Conference IEMC-2002, Cambridge 2002.

Tijssen, R.J.W., 2001. Global and Domestic Utilization of Industrial Relevant Science: Patent Citation Analysis of Science–Technology Interactions and Knowledge Flows. Research Policy 30, 35-54.

Trajtenberg, M., 1990. A Penny for your Quotes: Patent Citations and the Value of Inventions. RAND Journal of Economics 21, 172–187.

Trajtenberg, M., Henderson, R., Jaffe, A.B., 1997. University versus Corporate Patents: A Window on the Basicness of Invention. Economics of Innovation and New Technology 5, 19-50.

Tushman, M.L., Anderson, P.C., 1986. Technological Discontinuities and Organizational Environments. Administrative Science Quarterly 31, 439–465.

Tushman, M.L., Anderson, P.C., O'Reilly, C., 1997. Technology Cycles, Innovation Streams, and Ambidextrous Organizations: Organizational Renewal Through Innovation Streams and Strategic Change. In: Tushman, M.L., Anderson, P.C. (Eds.), Managing Strategic Innovation and Change. Oxford University Press, Oxford, 3-23.

Verspagen, B., 2000. The Role of Large Multinationals in the Dutch Technology Infrastructure. A Patent Citation Analysis. Scientometrics 47, 427-448.

Wasserman, S. & Faust, K., 1994. Social Network Analysis. Cambridge University Press, Cambridge.

www.european-patent-office.org/epo/pat_examiner.htm