Inventive progress measured by multI-stage patent citation analysis

von Wartburg, Iwan; Theichert, Torsten; Rost, Katja

DOI: https://doi.org/10.1016/j.respol.2005.08.001

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-68487

Originally published at:
DOI: https://doi.org/10.1016/j.respol.2005.08.001
Inventive Progress Measured by Multi-Stage Patent Citation Analysis

Iwan von Wartburg\textsuperscript{a*}, Thorsten Teichert\textsuperscript{b}, Katja Rost\textsuperscript{c}

\textsuperscript{a}University of Hamburg
\textsuperscript{b}University of Hamburg
\textsuperscript{c}University of Zürich

ZORA-Citation: von Wartburg, Iwan; Teichert, Torsten; Rost, Katja (2005). Inventive progress measured by multi-stage patent citation analysis. Research policy:1591-1607.

Abstract
Studies of technological change constitute a field of growing importance and sophistication. In this paper we contribute to the discussion with a methodological reflection and application of multi-stage patent citation analysis for the measurement of inventive progress. Investigating specific patterns of patent citation data, we conclude that single-stage citation analysis cannot reveal technological paths or lineages. Therefore, one should also make use of indirect citations and bibliographical coupling. To measure aspects of cumulative inventive progress, we develop a “shared specialization measure” of patent families. We relate this measure to an expert rating of the technological value added in the field of variable valve actuation for internal combustion engines. In sum, the study presents promising evidence for multi-stage patent citation analysis in order to explain aspects of technological change.

\textit{JEL classification:} O31

\textit{Keywords:} Patents; Technological Change; Bibliometrics; Citation Analysis; Social networks
Inventive Progress Measured by Multi-Stage Patent Citation Analysis

Abstract
Studies of technological change constitute a field of growing importance and sophistication. In this paper we contribute to the discussion with a methodological reflection and application of multi-stage patent citation analysis for the measurement of inventive progress. Investigating specific patterns of patent citation data, we conclude that single-stage citation analysis cannot reveal technological paths or lineages. Therefore, one should also make use of indirect citations and bibliographical coupling. To measure aspects of cumulative inventive progress, we develop a “shared specialization measure” of patent families. We relate this measure to an expert rating of the technological value added in the field of variable valve actuation for internal combustion engines. In sum, the study presents promising evidence for multi-stage patent citation analysis in order to explain aspects of technological change.

1. Introduction
Given that the study of technological change and knowledge generation is an important research field (Rosenberg, 1982, 1996), the question still remains of how such a complex, interrelated and cumulative process like technological change may be analyzed. The use of patents and patent statistics as indicators of technological progress stands for both, a long tradition and a controversial discussion about the value of patents as indicators of technological progress (Schmookler, 1966; Griliches, 1990).

Patents are a direct output category of industrial R&D and other inventive activity and mirror the cumulative process of technological change: on the one hand patent data enable longitudinal research and on the other hand they contain citation information that link different patents at different stages of technological development (Griliches, 1990).
cover almost every field of technology that is useful for analyzing the diffusion and the
development of key technologies. However, there are several drawbacks for patent data
analysis that have important implications for the validity of patents as measures (Pavitt, 1985;
Basberg, 1987; Griliches, 1990; Archibugi, 1992). Firstly and most importantly, not all
inventions are patented. Inventions may not meet the patentability criteria in the respective
jurisdiction, i.e. inventions have to be novel, non-trivial, and possess a commercial application
potential. The inventor might also strategically decide to exploit his intellectual property by
other measures such as keeping it secret or enforce copyright (Mansfield, 1986; Arundel and
Kabla 1998; Hall et al., 2001). Secondly, invention is not equal to 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.

The most common method for early patent data analysis was to simply count them and to
compare how many patents (applications and granted patents) had been assigned to different
entities, e.g. firms, nations, and technology fields. The validity of studies comparing different
technologies is dependent on the accuracy of the underlying patent classification systems. The
classification problem addresses the necessity and difficulty of mapping the technological
categories included in a Patent Classification Systems to economically meaningful entities
like products, firms or industries (Schmookler, 1966; Grupp, 1998). A second problem with
simple patent counts 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 sample mostly
accounts for the largest part of its value (Griliches, 1990; Harhoff et al., 2003). A simple
count of the number of patents equals these differences in value and importance out. In simple
patent count studies the phenomenon analyzed might at best be called “invention” and not
“innovation”.

3
The research progress achieved during little more than the last decade has been impressive. However, although today researchers can stand on the shoulders of giants there exists a shared understanding among experts that much work remains to be done in order to unpack the black box of technological change and knowledge generation. In this paper, we contribute to both, content and methodology in the study of technological invention. Our study aims to add to the discussion in three respects. Firstly, concerning the process of technological invention, we develop measures based on patent data to gauge the technical value added of inventions for a technological field in mechanical engineering. The technical value added of a single invention will be judged based on expert opinion in the form of technical scoring tables. The data set used to construct measures that help explaining technical value added consists of patent data. Secondly, we make use of network analysis and bibliometrics with patent citation data in order to develop an algorithm that allows laymen to cluster technical subfields. Thirdly, we contribute to the thriving discussion about patent analysis by looking at the explanatory value added researchers may gain by tracking technological change along patent citation chains longer than one citation step.

2. Theoretical development

2.1. Patent citation analysis

Patent citation analysis is based on the examination of citation links among different patents and between patents and scientific literature (Narin, 1994). 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 both by the inventor and the patent examiner with prior art in the respective technological field. The relevant sources for judging novelty and inventive step are referenced in the patent application.
A patent can be valuable from a technological (basicness) and/or from an economical (appropriability) point of view. The long-term \textit{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 subsequently. This relation is validated by a number of evidences (Carpenter et al., 1981; Albert et al., 1991; Harhoff et al., 2003; Trajtenberg, et al. 2002). The \textit{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. Empirical studies revealed a positive relationship between “times cited” of patents and firm performance or stock market valuations. Therefore, one may conclude that more often cited patents seem to generate higher economic value (Narin et al., 1987; Trajtenberg, 1990; Griliches, 1990; Deng et al., 1999; Breitzman and Thomas, 2002; Harhoff et al., 2003). Thus, by shifting the focus from counting patents to analyzing patent citation information, output success may be measured \textit{indirectly} and researchers may be legitimized to call the phenomenon studied not only “invention” but also “innovation”. However, addressing both, technological invention and commercial innovation may still be a too ambitious task. Since Schumpeter’s seminal work, these processes are believed to represent different phenomena (Schumpeter, 1939). Therefore, researchers who try to “get their arms around” both fields simultaneously may find themselves caught in a complexity trap. Thus, this paper will focus on technical invention. As theoretical underpinning we assume a path-dependent and boundedly rational search and selection process of technological inventions on behalf of participating organizations. Patent citation analyses base on "backward" measures (derived from the citations made by a patent) or "forward" measures (derived from the citations that a patent subsequently receives

\footnote{Citations between patents and scientific literature are not the subject of this paper, but they are important to mention. An interesting pioneering paper in which a lot of basic issues, also related to classification problems, are discussed is the work of van Vianen et al. (1990).}
from other patents). Forward citation studies used counts of patent citations as a measure of inventive quality in terms of technological or economic value (Harhoff et al., 1999, 2003; Henderson et al., 1998; Trajtenberg, 1990; Trajtenberg et al., 2002). Studies using backward citation information investigated spillovers called “knowledge flows” between technology classes (Trajtenberg et al., 2002; Rosenkopf and Nerkar 2001) or geographic regions (Jaffe et al., 1993; Tijssen, 2001). Other studies used backward citations as a means to analyze the localness of technological search by organizations (Podolny and Stuart 1995; Fleming and Sorenson 2001). It is important to state that the studies so far mostly used single-stage citation information. Yet, although Trajtenberg et al. (2002) use three generations of patents to assess the importance (forward citations) or basicness (backward citations) of inventions, their forward and backward measures primarily take into account the direct citations, i.e. the counts of indirect successors’ or predecessors’ citations is attenuated by a factor of 0.5. Thus, to investigate longer citation chains and getting more out of the historical citation information has rarely been undertaken.

2.2. On the comparability of citation data

Citation analysis has its origins within bibliometrics, i.e. the study of citation behaviour on behalf of scientific authors or academic journals. By analysing patent citation information, it has to be taken into account that there are important differences (Meyer, 2000). Firstly, the propensity to cite original references that mark the beginning of a thinking tradition is higher in scientific publications than in patent documents. In scientific literature, new articles often refer to classics in their field of study. Secondly, differences can be attributed to different motivations and actors in the citing process (Meyer, 2000). In scientific publications, references are stemming mainly from the authors and occasionally from comments on behalf of the reviewers. In patent documents, in any jurisdiction, the ultimate authority to place references on patent documents is the patent examiner. Compared to the sometimes flawed
citation behaviour of scholars (MacRoberts and MacRoberts, 1988), the use of patent citation data may be 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 the claims applied for by performing an in-depth review of existing patents and non-patent literature. The findings are documented as citations to patents and non-patent literature in the search and examination reports and eventually on the patent granted.

It is therefore questionable, whether the motivations for citing and the resulting citing behaviour of scholarly publications and patent documents can be compared. Meyer (2000) concludes that patent citation data adhere better to the citing ideal assumed by early bibliometricians such as Garfield (1972) than scholarly citation data. Meyer (2000: 112) also concludes that “one needs to have a thorough understanding of patenting practices in order to interpret patent citation data properly. Understanding the differences between European and US patent practice, for instance, is necessary to select the appropriate data sets for the analysis.”

2.3. On the compatibility of US and “European” patent citation data

An especially important caveat for patent citation analysis is the comparison of US with European or similar citation information. Firstly, search reports will end up containing different kinds of references for US and Europe. US patent law requires an inventor to disclose 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). Therefore, applicants provide extensive lists of citations in order not to run the risk of filing an incomplete list of references that may lead to the revocation of a granted patent in the future. Many times US examiners do not limit the lists of references provided by the applicants. In
the end, this leads to US search reports that have more of the characteristics of a documentary search than that of patentability search (Michel and Bettels, 2001). However, there is no duty of candour in Europe. Applicants can choose whether or not to supply a list of prior art. Even if they do supply one, examiners preparing the search report will screen and filter all the references according to their relevance with respect to patentability.

Secondly, examination procedures are different for US and European patent applications. Stated roughly, the examination procedures are less standardized in the US compared to the very standardized procedures in Europe. The distribution of the examiner at the USPTO cites per patents are highly skewed (Cockburn et al., 2003). The authors concluded that “examiners matter: Although highly structured, and carefully monitored by the USPTO, patent examination is not a mechanical process.” (Cockburn et al., 2003: 52).

Further, the patentability requirements are lower for US than for European patents (Cohen and Merrill, 2003). Quillen and Webster (2001) report that the patent approval rates in the USPTO are in the order of 80 to 90 percent or more. Therefore, in the US many references are making it to the front page of US patents that are not significantly relevant for patentability and one is able to find references counts in the order of more than 100 to 200 references (Harhoff et al. 1999, Hall et al. 2000). Besides, there is a strong “home bias”: US patent references tend to stay within the US patent system.

All these differences on behalf of inventors and examiners taken together lead us to conclude that US citation data seem to be less reliable for studying path dependencies in technological change due to a lesser degree of standardization in search and examination procedures. Concerning validity we do not want to take a stance for “European” or against US patent citation data. Keeping in mind that US citations tend to cite older patents than European patent documents, they have to be treated differently when deriving at variables based on the respective citation data. Therefore, to avoid biases in our study below we feel urged to solely
utilize citation data stemming from patents that were examined in a comparable way to the European search and examination procedure like German, British and PCT applications (WO patents) (Michel and Bettels, 2001). Given that the differences in US and “European style” citation data are now obvious it seems worthwhile to take a look at the guidelines and practices for examination at the EPO, in particular the search for prior art.

2.4. Examination at the EPO

Before the implementation of the “Bringing Examination and Search Together” project at the EPO, search and substantial examination were strictly separated tasks at the EPO. The search was carried out by dedicated search examiners at The Hague, whereas the substantial examination was done by a colleague in Munich. 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. 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 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. 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 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. Since there is no proof of completeness of a search report much emphasis is put on the training and experience of examiners. At several places in the guidelines, the examiner is asked to "exercise his judgment". The European patent office claims to have relatively low miss rates. This can be confirmed by additional,
much more profound searches that are carried out to identify novelty destroying subject matter.²

Today, the task of an examiner is to identify and to examine the closest state-of-the-art in patent and non-patent prior art in order to judge novelty, inventive step and commercial application potential of the claimed inventions. EPO examiners are asked to base their 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, the search will 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.

Then, the selection of documents is guided by the following principles (EPO Guidelines): 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, i.e. the respective patent from the to be cited patent 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

² The Swiss Federal Institute of Intellectual Property which supported this research project regularly performs such searches.
technically relevant information within a minimum number of citations.” (Michel and Bettels, 2001: 189)

The differences between patent citation and scholarly publication citation data mentioned above are rooted in this search procedure: because a minimum number of references is targeted cited patent references point only to the ‘closest’ prior art. Thus, single-stage citation entries regularly lack references to basic inventions in a technology field and are therefore not appropriate for the study of whole networks and lineages of technological inventions. Accordingly, to map actual developments in a certain technical field and to draw on technological trajectories (Dosi, 1982) or avenues (Sahal, 1985), citation analysis should rely on everything, bibliographical coupling, co-citations, direct and indirect citations. Based on this discussion, we infer three basic presumptions that should hold true when working on empirical studies with European or comparable patent citation data:

i. In contrast to bibliometrical studies where citation analysis has been originated, “European” patent citation data are unlikely to reference all relevant prior work.

ii. Citation analyses of “European” patents will reveal only loosely coupled direct cross citation networks among patents of a sample group.

iii. Extended “European” patent citation analysis that include multi-stage citation chains are able to reveal highly connected citation networks.

3. Measuring cumulative technological invention with multi-stage citation analysis

3.1. Network Measurements

A network for patent citation analysis consists of patent families and linkages between them (citing and cited citation information). Bibliometrical analysis draws on different information types (see Fig. 1). If a patent family cites or is cited by another patent family, this link is called a ‘direct citation’ link (B → D). If a cited patent family in turn cites another patent family, an ‘indirect citation’ chain between the first citing and the last cited patent family is in
place (B\(\rightarrow\)D\(\rightarrow\)C). A ‘bibliographical coupling’ refers to a ‘shared’ citation link where two or more patent families cite another other patent family (A/B\(\rightarrow\)D). A ‘co-citation link’ occurs if two or more patent families are cited together by another patent family (A\(\rightarrow\)C/D) (Braam et al, 1991a, 1991b; Egghe and Rousseau, 2002).

Fig. 1

Having a network of citations one is able to calculate network measures that characterize either holistic network characteristics or individual actors’ positions within a network. There are different approaches for evaluating individual actors’ positions within such a network (Wassermann and Faust, 1994). Three common definitions of centrality are degree centrality, closeness centrality and betweenness centrality (Freeman, 1979).

Measures of degree centrality have been implicitly deployed by previous studies and were used as an indicator of importance of inventions (Trajtenberg et al., 2002).

- Freeman in-degree: Importance in the sense of technological impact was calculated by counting the number of citations received and adding to that number the attenuated amount of citations received by the citing patents.

- Freeman out-degree: Importance in the sense of basicness of an invention was calculated by counting the number of references made by a patent and adding to that number the attenuated amount of citations made by the cited patents.

More basic patents possess low values for backward importance: “[…]our presumption is that more basic patents would have fewer and / or less important predecessors and therefore lower values of backward importance.” (Trajtenberg et al., 2002: 62). Thus, ceteris paribus, patents that are the offspring of a new technological trajectory are expected to score low on the freeman out-degree. Trajtenberg et al. found empirical evidence for this hypothesis using US patent citation data.
However, because we use citation information stemming from DE/EP/GB/WO patent families that are supposed to cite only the closest and most relevant prior art, we do not suppose that this finding will be reproduced. On the contrary, we expect that a simple count of backward citations for “European” style citation information will not yield a significant correlation between the amount of backward citations and the expert indicator for technological value added:

**Hypothesis 1.** The technological value added is independent from the number of backward citations of a patent family (Freeman out-degree).

Degree centrality measures, i.e. Freeman in-degree and out-degree may be refined along three dimensions for patent citation network analysis. Firstly, not all ties between patent families should be treated equally. Therefore, a probabilistic weighting scheme of ties between patent families is constructed and used for the measurement of direct and indirect citation links (section 3.2). Secondly, degree centrality measures only take into account the immediate ties of actors and neglect their indirect ties other actors. Thus, the considerations for measuring direct citation links need to be extended accordingly (section 3.3). Thirdly, patent citation information is of an inherent asymmetric nature, i.e. newer inventions citing older ones. Therefore, we have to consider bibliographical couplings and tie directional issues (section 3.4).

3.2. *Measurement of direct citation links*

To construct probabilistic measures of ties between patent families one has to take into account that citation links between patent families must not be of uniform weight and importance. We argue that the more prior inventions are referenced, the broader is the technological base of the citing invention. This holds since patent examiners are required to limit the cited references to the least amount required. 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. The rationale is that the new invention is likely to integrate certain aspects of both former ones and thus can be regarded as a hybrid development.

Thus, 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: if a citing patent references only a single other patent, a probability value of 1 will be 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 (see Fig. 2). This probabilistic interpretation limits the strength of direct relationships between any two patents to a maximum value of one.

3.3. *Measurement of indirect citation links*

Whereas in traditional network theory 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 scholarly publication citations in regard to the fact that references are strictly limited to the nearest, i.e. most recent prior art. 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 are captured by citation chains. In order to assess indirect citation relationships the considerations for direct citations are extended to longer citation chains:
• 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. A and B can be regarded as technological improvements of C. 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 (see upper part of Fig. 3).

• 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 (see lower part of Fig. 3).

Thus, we define the ‘reachability out-degree’ as the probability weighted direct Freeman out-degree times the probability weighted direct Freeman out-degrees of the cited patents. As the probability weight we use the inverse of the number of citations made. If a patent cites n patents, then every citation will be counted by a weight of 1/n. The reachability out-degree is a proxy for the “specialization” in what we call “shared specialization”. Thus, we hypothesize:

**Hypothesis 2.** The higher the reachability out-degree of a patent family (specialization), the higher the degree of technological value added.

### 3.4. Measurement of bibliographical coupling and directional changes

Patent citation data are 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. Given that patents A and D cite B we retrieve the directed relationships A->B as well as D->B (see Fig. 4). A strict vector interpretation of relationship patterns would suggest the absence of any relationship between A and D. Otherwise, neglecting the vector direction, one would infer
larger proximities for A-B and D-B compared to A-D (given by A-B-D). Both interpretations are inadequate: Since A and D built on B the shared amount of B-technology is the same for A and D. Thus, we can treat the bibliographical coupling A-D as being equal to the ties A-B and D-B.

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. This conceptualization of bibliographical coupling had been validated during discussions with experts of the Swiss Federal Institute of Intellectual Property. We expect it to be especially apt to fields of technology that show a pattern of cumulative innovation in the sense that new inventions will be more successful if they build on prior knowledge which is rooted in even earlier inventions and so on. Since a bibliographical coupling represents two patents citing a common third patent this is a proxy for what we call ‘shared-ness’ between these patents.

We define the degree of bibliographical coupling as one divided by the average of the direct out-degrees of the bibliographically coupled patents (see Fig. 5). Based on this discussion we infer hypothesis 3:

**Hypothesis 3.** The higher patent families are bibliographically coupled, the higher their technological value added (Shared-ness).

**3.5. Shared specialization measure**

In order to come to grips with patterns of specialized technological lineages and interrelated cumulative invention at the same time we have to combine the respective measures described.
Whereas the reachability out-degree measures technological lineages by extending the focus beyond direct linkages, the measure of bibliographical coupling is a proxy for the amount of “shared-ness” of technological features among technological variants. Thus, the former describes the specialisation along technological trajectories while the latter shows how much technologies have in common (shared technology). Together, the two measures can be viewed as a measure for “shared specialization”.

**Hypothesis 4.** The higher the patent families rate on the combined measure for ‘shared specialization’ derived by adding reachability out-degree and bibliographical coupling, the higher the degree of technological value added.

The expert criteria for application potential of patented technologies used in our empirical study (see section 4.1 below) include items like maintenance effort or manufacturing effort. These considerations are important from a managerial point of view for the evaluation of different technological principles to be incorporated in products. However, as we found in discussions with patent examiners, these criteria are not on the “hitlist” of concerns of examiners during patentability search and examination. Therefore, we do not expect any of our measures to be correlated with the expert opinion score on application potential of rated inventions:

**Hypothesis 5.** The degree of application potential is independent from the measures of Freeman out-degree, reachability out-degree, bibliographical coupling and shared specialization.

Likewise, we expect that the total score of the expert rating, i.e. the sum of the application potential, technological value added, and other criteria is unrelated to the proposed patent citation network measures:

**Hypothesis 6.** The total score of the inventions is independent from the measures of freeman out-degree, reachability out-degree, bibliographical coupling and shared specialization.
4. Methodology

4.1. Sample design and data gathering

We used the automotive sector to address the research questions and hypotheses. Data were collected from two sources: (1) expert ratings and (2) archival data from patent records. By choosing an indicator for technological impact that is based in expert judgement of a technology we follow Albert et al. (1991) who used expert opinions in order to directly validate the measure of received citations of patents as a proxy for technological quality.

*Expert ratings.* The choice of our core patent sample was driven by an engineering dissertation in the field of four strokes internal combustion engines. Hannibal (1993) studied 201 camshaft and variable valve actuation patents and valued their quality along 14 criteria. In his sample design, he focused on camshaft and valve control patents that at the time of his dissertation’s publication seemed to be the most promising ones in terms of their potential to influence the makers of vehicles in the next years or decades. Variable valve actuation denotes the mechanism used to shift the cams to optimal positions depending on RPM. A fixed-profile camshaft will be optimal only at one engine speed. At every other engine speed, the engine won't perform to its full potential. This is why carmakers have developed schemes to vary the cam profile as the engine speed changes.

Table 1 contains our division of the criteria chosen by Hannibal (1993) into indicators of technological value added and application potential. The former indicates the inherent technological value of an invention. The latter indicates the potential that the invention may become an *innovation* by becoming incorporated in engines which will be manufactured and built into cars. The 14 criteria together make up the total score of a valued patent family. These indicators still limit our conjectures to the level of inventions but they allow for also assessing the application *potential* of inventions in terms of manufacturing or maintenance efforts required and the like.
Table 1

Archival data from patent records. Taking the patents valued by Hannibal (1993) as our core sample we gathered a snowball sample. A snowball sample begins with a set of actors (Goodman, 1961). The actors to whom they have ties of a specific kind constitute the first-order zone. Actors tied by the actors in the first-order zone who are not among the original respondents constitute the second-order zone. This snowballing proceeds through several zones. In this study we walked “backward” by gathering patent documents referenced by the core sample. We proceeded as follows:

- First we searched for the 201 Hannibal-patents in EPOQUE’s EPODOC Database and found 178 patents. Almost all missing patents were Japanese patents and applications.
- We then grouped the patents found by patent families. Only 4 patent families contained 2 of the original patents, so that in our core sample 174 core patent families remained.
- Next, we gathered all the family members of the 174 families, i.e. 671 patents. We then selected this group of patents by applying what we call the “DE/EP/GB/WO”-Filter. It selected only the German (DE), European (EP), British (GB) and World (WO) patents out of the family members of the core sample patent families.
- The remaining 276 DE/EP/GB/WO family members were taken as the starting point for extracting the first generation of references made by the core family members. The 276 family members cited 403 different patents. Out of these, 385 patents were found in and downloaded from the EPODOC database.
- These 385 patents were associated with 981 family members. From all these family members, we again selected the DE/EP/GB/WOs, i.e. 347 patents. Together with the Non-DE/EP/GB/WO’s out of the 385, 540 patents that were referenced by the core patents entered the database. 71 patents were already contained in the core patents. That means that there are cross citations within the core patent sample.
• The 347 DE/EP/GB/WO’s served as the starting point for extraction of the second generation of references made. The 347 family members cited 503 different patents. Out of these 489 were found in EPODOC and entered in the database. The 489 then were grouped by patent families and again the respective family members were collected: 1333 patents. The amount of DE/EP/GB/WO’s is 463. From these 463, 191 did not enter our database priorly and were entered. Since we did not go any further backwards in looking for references made, our reference hunting routine came to a halt.

4.2. Development of the citation network

As step one we looked at the cross citations among the core sample patent families and found only 53 patent families that were in some way connected to other of the 173 nodes in the core sample. Thus, by drawing this citation network we find a very small and loosely coupled network. As a second step we traced citation chains of the length of two: core patent families cite a non-core patent family which in turn cites a core-patent family. By doing so, we were able to add 21 patent families to the small citation network found in step one. After that we added citation chains of length three, i.e. we ‘walked along’ three edges: from a core sample patent family to a non-core family (first step). From there to another non-core family (second step) and finally back to a core family (third step). By doing so, we arrived at 117 connected patent families of the 174 core patent families (see Fig. 6).

![Fig. 6](image-url)

From this network consisting of 117 connected patent families, we extracted the largest network component. A component is a part of a network that is fully connected. It contains 107 core patent families. By inserting attribute data for different groups of technological solutions for variable valve actuation proposed by Hannibal (1993) we arrived at Fig. 7. The technological clusters in the main component are approximately marked by circles. Each circle represents a different technical principle to achieve a greater variability in valve
actuation. Core-patent families are marked by rectangles. Dots symbolize non-core patent families found during the first and second citation steps which lead back to core-patent families. Since the patent family grouping algorithm in the programmed database system takes the first patent found in a patent family as its label, it is possible that a patent family containing at least one DE/EP/GB/WO patent carries a US-label.

From analyzing this derived main citation network component we can already observe some interesting preliminary findings. Firstly, the resulting clusters by analyzing patent citation chains by the length of three match the clusters identified by the expert opinion pretty well. We interpret this as both a positive sign for the applicability of patent citation analysis along longer citation chains for European patent citation data and for the validity of the expert judgment we are relying upon for the technological impact measure in our study. Secondly, the patterns found are in accordance with the discussion of the “European” examination practice. If the patents in our core sample had cited significantly more, our cross citation network resulting from step one would already be denser and we would not observe such high levels in denseness increase by walking along the citation chain. Thus, we find support for the proposed basic presumptions. Thirdly, one can infer that variable valve actuation indeed represents a technological field which is characterized by highly cumulative and interrelated inventive progress. We arrive at this conclusion by observing that the removal of certain technological clusters leads to a falling apart of the main component of the core patent family citation network. By removing one of these binding technological clusters, we are able to reveal which technological clusters draw on the removed cluster. These clusters can be considered to represent different offsprings of an initial technological avenue. They recombine what has been proposed by the initial technological variant. Accordingly, the
technical field of variable valve actuation seems to be appropriate for deploying the measures for “shared specialization”.

4.3. Calculation of measurements

In order to run the measurement calculations we relied on the whole citation network containing core patent families as well as non core patent families. In fact, all the patent families that were stored in the database during the data gathering process described above were used as sources for backward citations. These citations were extracted by appropriate queries using SQL language and converted into an edge-list format in order to represent input data for the network analytical software package UCINET 6.0. After creating the whole citation network we calculated the freeman out-degree, the reachability out-degree and the bibliographical coupling measures. In order to calculate the indirect citation measure and the measure for bibliographical coupling, several transformations of the initial adjacency matrix representing citation information were necessary.³

5. Results

5.1. General results for the shared specialization measure

By applying the algorithms of the proposed network measures, i.e. direct citation weights, reachability out degree, bibliographical coupling, and the combined measure for “shared specialization” we derive at a new network visualized in Fig. 8. The network now represents the proximity of technological clusters. Previously unrelated patent families are now related via direct, indirect, and bibliographical coupling linkages.

---

³ A detailed description of the algorithms and the matrix algebra used can be requested by contacting the corresponding author.
From a comparison of Fig. 7 and Fig. 8 we can infer that the previously identified bridging technology clusters still bridge the previously highlighted technologies: The cluster “variability of camshaft drive” still links together the different subgroups of the cluster “variability through different linking mechanisms of cams and valves”, i.e. the hydraulic and mechanic subcategories. Further it is evident, that the technological cluster which seeks variability through an additional camshaft is occupying central positions linking the whole closeness network together. This makes intuitively sense, since including additional camshafts is an opportunity that can be followed in the most of the already described technological clusters. The cluster “variability by electrical systems without camshafts” represents a technological trajectory that has been cited more from other clusters that it cites them. This observation can also be readily explained, since this the cluster without a camshaft is the only “camshaft-free” variable valve actuation solution that draws necessarily on certain functioning principles of (prior) solutions employing camshafts.

5.2. Hypotheses proof

To investigate the proposed hypotheses the calculated measures were correlated with the variables taken from the expert rating. Table 2 shows the resulting correlation matrix. We find no significant relation between the Freeman out-degree measure and the technological value added of the citing patent families determined by the expert score. Thus, we gain preliminary support for hypothesis 1. By including indirect citation links and measuring the logarithm of the reachability out-degree measure we find a significantly positive correlation with the technological value added. Thus, we also gain support for hypothesis 2. By looking at bibliographical couplings as a measure of bringing different technical principles together, we find a significantly positive correlation of the bibliographical coupling of core patent families with technological value added. Hypothesis 3 is supported. The same holds for hypothesis 4:
the measure for shared specialization is significantly positive correlated to technological value added.

However, our network measures do not correlate significantly with the expert rating on application potential of the inventions. Thus, hypothesis 5 is supported. Hence, no matter how sophisticatedly one deploys patent data, if one uses them to judge some sort of an economic value of patents, they will embody strongly limited indicators in cases where this value is rooted in cost efficiencies concerning the operational sphere of companies. This is an important caveat when working with patent data in inventive fields that are characterized by cumulative and interrelated technological change.

Table 2

Since both, bibliographical coupling and shared specialization are significantly and positive correlated to the total score of the expert rating we find no support for hypothesis 6. Although the total score correlates heavily with the valuation of the application potential of inventions, bibliographical coupling and shared specialization measures are significantly related to the total score and not related to the application potential. Given that the correlation between bibliographical coupling and total score is stronger than the relation between shared specialization and total score we may offer a tentative explanation for this finding: the cumulative nature and interrelatedness of technological change in variable valve actuation places a premium on the combinatory search when designing new inventions. However, this proposition needs dedicated further research.

Conclusions

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. With the presented study we contributed to the discussion with an in-depth methodological reflection on the potential of patent citation network analysis for explaining technological change. Pointing to the very specific patterns of patent citation information, we concluded 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 explained that standard measures of network analysis have to be adjusted to the specific conditions and purposes of patent citation network analysis. Most notably the degree centrality measure (Freeman out-degree) has to be questioned as an indicator of technological basicness of inventions. This is especially true if one uses “European style” patent citation data and is less true if working with US patent citation information (Trajtenberg et al., 2002). When using mixed datasets consisting of “European” and US patents one should be cautious because the analysis can become biased because of the home-bias in citations or the significantly higher number of citations on US documents. As a consequence, we introduced a novel approach for assessing technological proximity based on such adjusted measures in patent citation networks. We specified rules for measuring both direct and indirect citation links and propose a way to deal with bibliographical couplings. In particular, in directed citation networks there is no connection between later patents that share prior technology. The measure of bibliographical coupling developed in this paper adds to the current discussion of patent citation networks.

To exemplify the feasibility of our approach, we relied our patent citation network measures to expert evaluation of technological value added of the patent families. These measures included the reachability out-degree as a proxy for the specialization in technological lineages and the bibliographical coupling as a proxy for the shared-ness of priory known technological
principles in inventions. We found that inventions with a higher technological value added relate significantly positive to the measures of reachability out degree, bibliographical coupling and the additive index of shared specialization. The Freeman out-degree does not relate to the technological value added of inventions. If these measures can be reproduced in future studies in different fields of technology characterized by supposedly different patterns of inventive progress (e.g. pharmaceuticals) it can become a viable complement to other proxies of patent quality like amount of citations received, geographical breadth, amount of claims or duration of renewal fees payment (Lanjouw et al., 1998).

From a managerial perspective this would be a worthy achievement: Except for the amount of claims, the other proxies are only retrievable after a certain amount of time. However, the measure of shared specialization can be determined by using backward citation information that is available ex ante. Therefore, from a technology intelligence point of view, further validation of a measure for shared specialization could become a promising avenue for further research. Another achievement of the presented approach with managerial relevancy is the possibility to classify technological subgroups efficiently. Again, before claiming a universal viable method for grouping technological fields in a cost and time efficient manner, further validation examples are necessary.

Last, we like to point to several deliberately chosen simplifications and omissions in our study that can be relaxed or included in the future. Firstly, we focus on one citation information in patent documents: patent citations. We do not further consider non-patent-citations like scientific literature. By adding this information, one could perform methodologically novel research about the science-technology linkage. Secondly, in this study the weighting of patent citation relationships between patent families was done in a probabilistic manner. For “European” citation information there is precious additional information in patent documents that was not included yet in this study: the references made by examiners are classified
according to their specific relevancy for the patentability decision. There are, for instance, “particularly relevant documents” marked by an [X] taking away the novelty of the essential features of the invention, “documents describing the technical background” marked by an [A] or documents that are particularly relevant in combination with other documents marked by an [Y]. One could use this weighting information in order to enhance the validity of the constructed measures. Thirdly, this study looks at inventive progress exclusively from a technical perspective. We may gain many more interesting research findings when including the relational context of the inventors as this has been done for example by Podolny and Stuart (1995) or by Balconi, Breschi, and Lissoni (2004). Fourthly, we are aware of the fact that methodologically much more can be done in terms of multivariate analysis of network calculations. Keeping the limitations of the study explicitly in mind, we nevertheless see preliminary evidence for great potential benefits of patent citation network analysis and thus encourage further research in this area.
Acknowledgments

We thank Harald Pier and his colleagues at the Swiss Federal Institute of Technology for supporting this research with expert insight concerning search and examination procedures.

We also thank a peer reviewer for thoughtful comments concerning the state of the art in bibliometrical analysis.

References


Figures

Fig. 1. Citation types.

- A is a direct successor of B, Proximity = 1
- A encompasses both aspects of B and C, Proximity of A to B and A to C = 0.5 or –in general– 1/n (n cited patents)

Fig. 2. Measurement of direct citation links.

Fig. 3. Measurement of indirect citation links.
Fig. 4. Interpretation of directional changes and bibliographical couplings.

Fig. 5. Measurement of bibliographical couplings.

A \rightarrow B \rightarrow C \rightarrow D \rightarrow E

P=\frac{(1/n)+(1/m)}{2}
Fig. 6. Stepwise development of the citation network.

Fig. 7. Main component of citation step two network with non-core families grouped according to technological clusters.
Fig. 8. Patent citation network after running the closeness algorithms
Table 1
The criteria for the evaluation of the variable valve control patents

<table>
<thead>
<tr>
<th>Dimensionality</th>
<th>Criterion</th>
<th>What it measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technological value added</td>
<td>Functionality</td>
<td>Detailed valuation of the basic and general technical features. Central for development of products. Low rating is only possible for research prototypes.</td>
</tr>
<tr>
<td>Variability</td>
<td>Variability</td>
<td>Variability of valve lift curve. There are different avenues to improve on the valve lift curve that are incommensurable, i.e. cannot be combined within one solution.</td>
</tr>
<tr>
<td>Application potential</td>
<td>Development effort</td>
<td>Effort for the development of the technical into a productized solution. Dependent on the degree of variability: more variability means more development effort.</td>
</tr>
<tr>
<td></td>
<td>Space</td>
<td>Need of space for the variable valve actuation in a motor</td>
</tr>
<tr>
<td></td>
<td>Construction effort</td>
<td>Effort to build the solution in question</td>
</tr>
<tr>
<td></td>
<td>Information</td>
<td>Amount of additional electronic information needed in order to steer the variable valve actuation</td>
</tr>
<tr>
<td></td>
<td>Cost</td>
<td>Need for capital and expenditures in order to construct and develop a solution</td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>Complexity, availability and flexibility of process to manufacture the solution in masses</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Feasibility and ease of quality control</td>
</tr>
<tr>
<td></td>
<td>Maintenance effort</td>
<td>Effort and time interval for maintenance</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Possibility to influence the time needed to develop and commercialize a solution for variable valve actuation</td>
</tr>
<tr>
<td>Other</td>
<td>Energy efficiency</td>
<td>How efficiently energy is used in order to improve the energy effectiveness of the motor.</td>
</tr>
<tr>
<td></td>
<td>Security</td>
<td>Risk for the driver in case of a defect of variable valve control</td>
</tr>
<tr>
<td></td>
<td>Environment</td>
<td>Amount of emissions for the environment (noise, gasoline, etc.)</td>
</tr>
</tbody>
</table>

Table 2
Resulting correlation matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Technological Value Added</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Application Potential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Total Score</td>
<td>.48**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Freeman Out-Degree</td>
<td>-.09</td>
<td>-.02</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Reachability Out-Degree</td>
<td>.38**</td>
<td>.14</td>
<td>.14</td>
<td>-.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Bibliographical Coupling</td>
<td>.34**</td>
<td>.08</td>
<td>.28**</td>
<td>.11</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>7 Shared Specialization</td>
<td>.45**</td>
<td>.08</td>
<td>.27*</td>
<td>.03</td>
<td>.53**</td>
<td>.93**</td>
</tr>
</tbody>
</table>

n=107, * p<.05, **p<.01