Scientometrics, Vol. 30, No. 1 (1994) 147-155

PATENT BIBLIOMETRICS*

F. NARIN

CHI Research, Inc., 10 White Horse Pike, Haddon Heights, New Jersey 08035 (USA)

(Received January 18, 1994)

In our 1975 monograph "Evaluative Bibliometrics" we discussed the many uses of publication and citation analysis in the evaluation of scientific activities, and some of the basic statistical properties of the scientific literature, particularly the skewness of the distributions of publications and citations, reference time distributions, and various anomalies in the citation patterns from one country to another. Over the last ten years we have devoted much of our energy to the development of an analogous research base and infrastructure for patent bibliometrics, that is for the use of patents, and patent citations in the evaluation of technological activities. There are remarkable similarities between literature bibliometrics and patent bibliometrics, and they are both applicable to the same wide ranges of problems. This paper will show that there are striking similarities between literature and patent distributions of national productivity, inventor productivity, referencing cycles, citation impact and within country citation preferences.

National productivity

Most scientists and scholars are much more familiar with the research literature and techniques of bibliometrics as applied to science and scientific papers, than they are to the applications of the same techniques to technology and patents. In this paper we will review some of the macro-scale indicators that have been developed for both patents and for papers, and show that there are far more similarities than differences between these two realms.

It should be mentioned at the outset that the idea of using counts of the published products of scientific research to measure the activities in that realm is not new. The first paper of which we are aware is a beautiful paper by *Cole* and *Eales* published in 1917 tracing the history of comparative anatomy from the years 1550 through 1850.[1] In their paper *Cole* and *Eales* tackled many of the problems that are still important in the analysis of science and technology: whether a paper or book or patent should be attributed to the country of domicile or the country of origin of the scientist, how to

* Paper presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics in Berlin (Germany), September 11 – 15, 1993.

Elsevier, Amsterdam – Oxford – New York – Tokyo
Akadémiai Kiadó, Budapest

differentiate between a single publication or patent of great impact and many publications of more routine interest, and so forth, and how to deal with the military and political events which have clearly effected the patterns of publication and patenting.

The first person to place the amount of scientific publication into an economic context was *Derek De Solla Price*, who first showed that the amount of scientific publication coming from a country was essentially proportional to its economic size, as measured by its gross domestic product, and not proportional to geographic area, population or any other parameters.[2] Price's point was that a research publication was a characteristic of a modern nation, and proportional to the economic size of the nation.

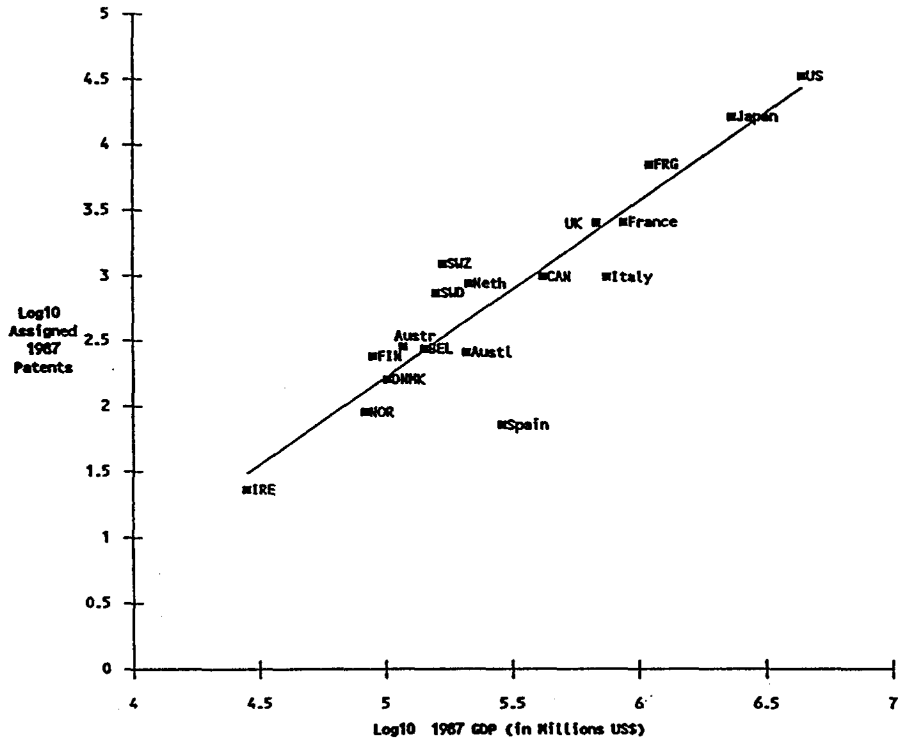

Fig. 1. Number of assigned U. S. Patents *vs* GDP for 18 countries

That work has been generalized in a number of subsequent papers, and has been found by us to hold not only for scientific publication, but also for patenting. This is shown in Fig. 1, which shows the number of patents granted in the United States patent system to inventors from different countries, plotted against GDP. Note that the association is quite striking, and its clear that both the scientific and technological productivity of countries, as measured by papers and patents, are to a very close approximation associated primarily with economic activity.

Inventor productivity

Understanding the role that key men and women play in human affairs has always been a fascinating aspect of human achievement. This was first approached from a quantitative viewpoint toward the end of the last century, when Francis Galton developed various measures of eminence such as inclusions in biographical compilations or in selected columns of obituary notices. He noticed that these indicators were very highly skewed, with a small number of very eminent people and a much larger number of people of more ordinary accomplishments.[3]

In the realm of scientific productivity this was first quantified in 1926 in what has now become known as 'Lotka's Law' which is a $1/n^2$ productivity law.[4] *Lotka* first studied the number of papers attributed to specific scientists in chemical abstracts, and noticed that they were distributed in the following way. For each 100 scientists who publish one paper there were approximately $100/2^2$ or 25 who published two papers, $100/3^2$ or 11 who published three papers, etc.

There have since been hundreds of papers published showing that Lotka's Law holds for many collections of scientific papers.

The first demonstration of this type of concentration for patented technology was shown in a paper of CHI's which is currently in publication.[5] In that paper, semiconductor technology patents of Xerox, AT & T, Fuji Electric, and Matsushita Electric were analyzed.

Figure 2, the Xerox Inventor Stick Diagram, is an illustration of the data found. Note that 119 Xerox inventors received only one semiconductor patent in the eight years covered, 16 received two, 10 received three, and so forth, with one very prolific inventor whose name is on 18 different patents.

This is a Lotka type of distribution, and is also found in the other three companies, although the distribution in patents may be somewhat steeper ($1/n^k$, $k > 2$) than in papers. But the general principal, of concentration of productivity in a small number of researchers, clearly holds both for literature and for patent bibliometrics.

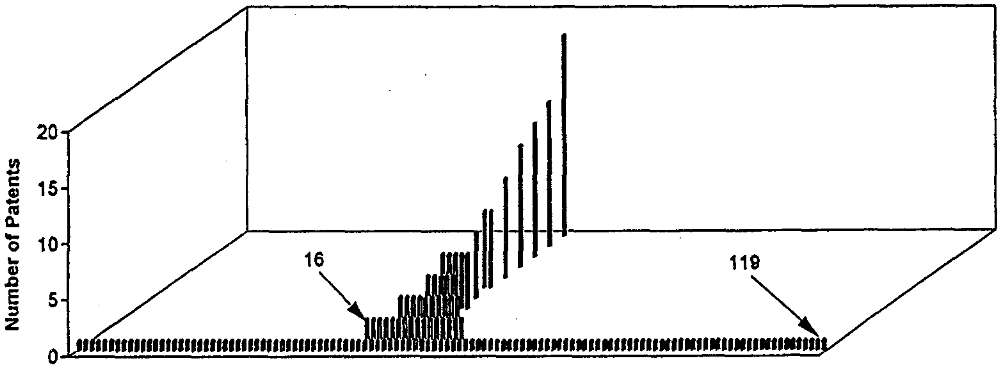

Fig. 2. Xerox inventor productivity stick diagram: Semiconductor patents 1984–1991. Each stick represents 1 inventor and all his patents (whole counts)

Referencing cycles

The old "linear model" of the way in which science, technology, and economics develop, postulates that scientists do research, publish, and place that information in the public domain. It then gets codified, taught, and eventually works its way into technology and is utilized by inventors who produce patented inventions, which are manufactured and sold, with a consequent contribution to economics. This linear model, science, then technology, then economics is simplistic and clearly inaccurate, and ignores the much more intimate relationship between technology and science, with many feedback loops.

The key additional point we wish to make is that the implicit, step-by-step time scale of this model is also wrong. In some of the modern areas of technology there is now virtually no time lag between science and technology. In an earlier *Scientometrics* paper entitled "Is Technology Becoming Science?" we showed, in essence, that the time lags, as measured in referencing cycles, in biological areas, are extremely similar for scientific papers citing earlier papers, for patents citing earlier patents, and for patents citing earlier scientific papers.[6]

What this says, in essence, is that the two realms of science and technology are very closely linked. In biotechnology there is essentially no time lag between science and technology. The inventor works in the university or the government lab in the morning in the United States, and he works at or consults with a private company in the afternoon, and the time lag between his academic research and his private inventive activity is lunch!

That is not to say that there aren't significant differences in scientific and technology cycle times, both from country to country and from field to field.

For example, Biochemistry and Molecular Biology papers have two to three times the number of references/papers as acoustics papers, and the references are to much more recent papers (shorter cycle time).

There are similar, large differences in cycle times between individual areas of technology. If we define Technology Cycle Time (TCT) as the median age of the U. S. patents referenced on the front pages of other U. S. patents, then Electronics, a relatively fast moving area, has cycle times of four to five years in the US patent system, whereas Ship and Boat Building and some of the older mechanical areas have cycle times that are more likely to be in the range of 15 to 20 years.[7] Again, a striking analogy between scientific and patent bibliometrics, and even more striking relationships between them.

Citation impact and distribution

Another area where there are striking similarities between scientific and technological realms is in the distributional properties and interpretation of citations. It is now quite widely accepted that the frequency with which a set of 25–50 or more papers is cited is a good indication of the importance and impact of those papers. This has been established in hundreds of papers for sets of important scientists, for areas of research, for departments, for universities and laboratories, and so forth.

It is perhaps not as widely known that the same phenomenon holds for US patents. This has been shown to be valid for patents in two specific papers, a very early paper of ours which showed this in general, and a much more recent one which showed this to be true in an industrial environment.

In the first validation paper we showed that patents associated with The Industrial Research Institute IR100 awards were almost twice as highly cited as a random set of patents picked from the same areas of technology.[8] While we did not know that each specific patent was associated with that specific invention, we did pick patents by the same company and by the same inventor, in the area of the IR100 award, and found that these were much more highly cited than a comparable set of random patents.

In a more recent study, we directly compared citation frequency of a set of patents in Silver Halide technology, invented by inventors at the Kodak Company, with senior Kodak staff opinion of the technological importance of the advances contained in those patents.[9] We found a strong correlation. In particular, in the Kodak study, it was quite clear that the patents which had received a relatively large number of citations, five or ten or more, were much more highly ranked by the peers than patents with a relatively small number of citations.

We had also observed in the IR100 awards study, that there was a rather steep distribution, with a relatively small number of patents receiving many citations, and the majority being very lightly cited, if cited at all.

This skewness of the citation distribution is a common characteristic of papers and patents. In both of these realms there are a relatively large number of papers or patents which are not very highly cited, and a relatively small number that are highly cited. Since the US patent system consists of 100,000 patents a year, each of which contains five or six references on its front page, patent citation counts are much smaller than citation counts for the *Science Citation Index* where, for example, there are some 500,000 papers a year, each of which contains ten or more references to earlier papers. Nevertheless, in both realms the distributions are highly skewed, again illustrating the remarkable similarities of the two areas of bibliometrics.

Within country preference

The last point we would like make in this paper deals with the very strong within country preferences shown both in the citation patterns of research papers, and in the citation patterns linking patents to the scientific literature.

With respect to papers citing papers, Fig. 3 shows the pattern of citation from U. K. authored papers in clinical medicine to earlier U. K. authored papers in clinical medicine. Note that the immediately cited papers, and the very recent papers cited are highly likely to be U. K., 50 to 70 percent, whereas the older papers cited, cited papers five or more years old, are only 35 to 40 percent U. K. This citation time anomaly is characteristic of the citation patterns for every country and every field: there is very quick and heavy citation to a country's own papers, and much slower citation to papers outside of the country.

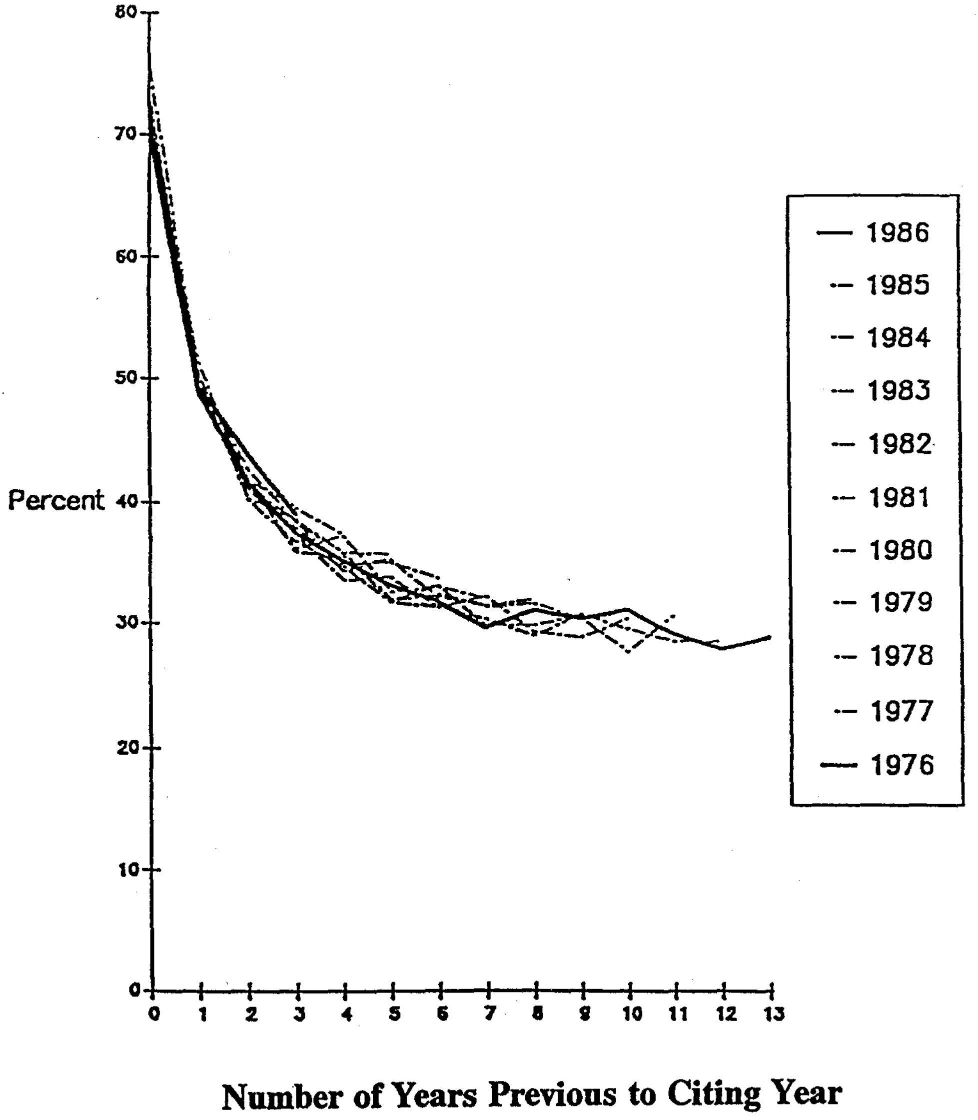

Fig. 3. Citation time anomaly. Percent of U. K. cites to U. K. papers. 1973 journal set – clinical medicine

Obviously, a large fraction of this has to do with a scientist citing his own papers, and the papers of his colleagues at his own university, with which he is very familiar, often before they are actually published. So this is really an indicator of the local nature of science, and the dependence of one persons' work on his own and his close colleagues earlier work. Nevertheless it is a very important and very strong indicator, and has a major effect on any cross national citation comparison. For example, in a field such as Biomedicine, in a database which is so heavily US dominated as the SCI is, with close to 40 percent US biomedical papers, the citation to US papers occurring in the first two or three years will be heavily biased toward the US, and therefore exaggerated compared to the long term citations to papers coming from the rest of the world. Similarly, immediate citation patterns to papers from smaller countries will be much lighter than the long term ones, when compared to the larger countries.

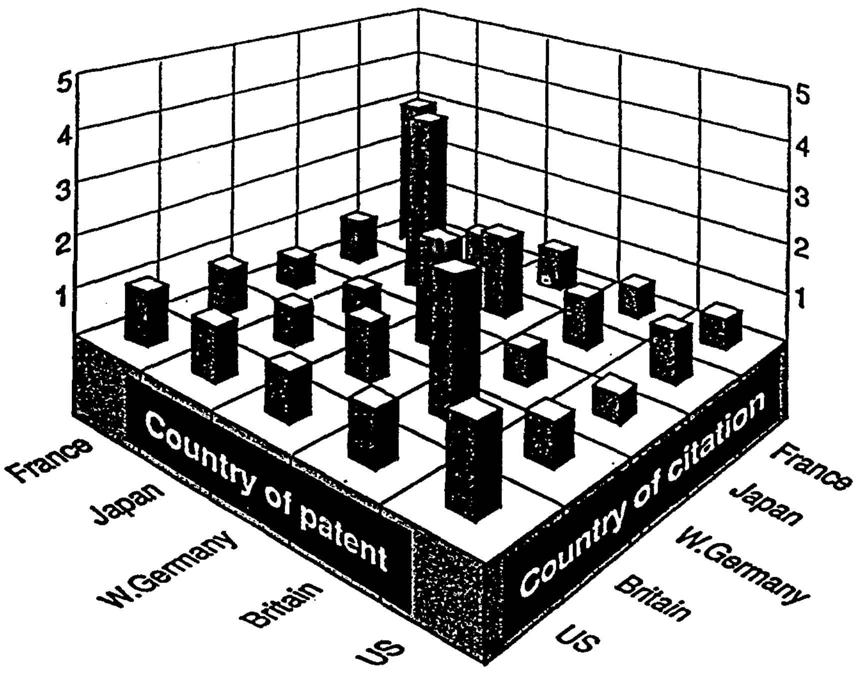

Fig. 4. Linkage between technology and science

In the final figure, Fig. 4, we show citations to scientific papers from the front pages of US patents, on a citing and cited country basis, that is, citations from each country's inventors to each other country's research papers. The data are fully normalized: that is, citation percentages are divided by publication percentages for the country's papers, so that the expected values are all 1.0. Thus, if 18 percent of the science citations from German patents go to German papers, and if six percent of the scientific papers are German, then the normalized ratio would be 3, approximately as shown on the figure. For every country we see essentially a local domestic citation anomaly factor of 2 or 3, linking technology to domestic science.

Thus, both in the scientific and technological realms there are again similar patterns of national linkage, and again a good deal of similarity between literature and patent bibliometrics.

We thus conclude, as *Price* observed many years ago, that there are more similarities than differences between patents and papers, and technology and science, and that the general bibliometric properties of these two realms are very similar.

References

1. F. J. COLE, N. B. EALES, The history of comparative anatomy, *Science Progress*, 11 (1917) 578–596.
2. D. J. DE SOLLA PRICE, Measuring the size of science, *Proceedings of the Israel Academy of Science and Humanities*, 10–11 (1969).
3. D. J. DE SOLLA PRICE, *Little Science, Big Science*, New Haven, Yale University Press, 1963.
4. A. J. LOTKA, The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, 16 (June 1926) 317–323.
5. F. NARIN, A. BREITZMAN, Inventive productivity, to be published in *Research Policy*, December 1993.
6. F. NARIN, E. NOMA, Is technology becoming science? *Scientometrics*, 7 (1985) 369–381.
7. F. NARIN, *TECH-LINE^{TM} Source Book.* CHI Research, Inc. May 1991.
8. M. P. CARPENTER, F. NARIN, P. WOOLF, Citation rates to technologically important patents, *World Patent Information*, 3, 4 (1981) 160–163.
9. M. B. ALBERT, D. AVERY, P. MCALLISTER, F. NARIN, Direct validation of citation counts as indicators of industrially important patents, *Research Policy*, 20 (1991) 251–259.