

# Current Comments®

The 1955-1964 *Science Citation Index*  
Cumulation—A Major New Bibliographic Tool  
for Historians of Science and All Others Who  
Need Precise Information Retrieval for the  
Age of Space and Molecular Biology

Number 5

January 31, 1983

In the epilogue of my book *Citation Indexing—Its Theory and Application in Science, Technology, and Humanities*,<sup>1</sup> I expressed the hope that the retrospective coverage of *Science Citation Index*® (*SCI*®) would one day go back to the beginning of this century. This extension of *SCI*, I believe, will be invaluable for historians and scientists in studying the evolution of modern science. Sometime in early 1984, ISI® will publish a ten-year cumulation of *SCI* covering the scientific literature from 1955 to 1964. This represents a giant step toward the ultimate goal of completing our coverage for this century.

It is now more than 20 years since the appearance of the first printed *SCI*. Then, as now, *SCI* exploited the cognitive links between ideas, which citations represent, to provide a multidisciplinary tool for searching the literature. But from the start, it was apparent that the *SCI* data base would serve for more than just information retrieval. Over the years, we have used the data in *SCI* to identify high impact papers, journals, and authors. More recently, we've applied citation data, through our technique of co-citation clustering,<sup>2</sup> to mapping the structure of science. This application of automatic classification combined with human scientific editing has already been used in our recently published prototype *ISI Atlas of Science*™.<sup>3</sup>

To date, the 1961 volumes have been the earliest *SCI* available. The new ten-year cumulation will extend this coverage back six years. It will cover a decade which is already of considerable interest

to historians. Consider that during the decade 1955-1964, the space age began, as did molecular biology and countless new fields too numerous to mention here.

There is a considerable body of historical literature now available which has been stimulated by the availability of *SCI*. The *SCI* has made it possible to study developments in specific fields in a way that was almost impossible before. Of course, historians used traditional indexes in the past to study the reception of scientific ideas by contemporary researchers. But the unique ability of *SCI* to tell you exactly when and where a particular paper or author is first cited or discussed is an important ingredient in its widespread use. Sometimes a simple glance at the printed *Citation Index* will give you an idea of the degree of receptivity of a particular idea at the time it was first reported. This has little to do with the as yet controversial notion that the quantity of citation may reflect the relative importance of different scientific events.

The 1955-1964 cumulation will not be limited in interest to historians of science. The fact is that almost ten percent of the citations in any recent *SCI* annual cumulation are to material published from 1955 to 1964. You can often find current references to these older articles in a recent *SCI*. But if you want the full title of the paper in question, you must go back to the original in order to complete your citation. Since librarians are often given the task of completing or verifying such citations, they too have

long appreciated the unique value of *SCIs* cumulated *Source Index*. Among other reasons, they need not decide if the subject falls into the province of a particular discipline-oriented index. And without the 1955-1964 data, you may never be able to determine what the impact was of a particular paper of this period at the time it was published.

It is difficult to say exactly when we began to plan the 1955-1964 *SCI* cumulation. Such a project, of course, had long been a dream of mine. During our *Genetics Citation Index* project,<sup>4</sup> we compiled data covering those years for a small group of genetics journals. This is how we demonstrated the need for multidisciplinary coverage. But it wasn't until 1979 that we actually began to acquire back issues of the more than 500 journals we identified as core literature for the decade covered. These journals were identified through citation analysis, and cover virtually every active specialty of science at that time. In addition to complete coverage of the core 500 journals, the cumulation includes all the journals covered in the corresponding 1961-1964 annual cumulations of *SCI*. But in order to make our coverage complete for the decade, we "filled in" data for a number of journals not yet covered in *SCI* in the period from 1961 to 1964.

The new *SCI* cumulation will contain a *Source Index* and a *Citation Index*. In order to reduce the cost of processing such a large volume of back issue material, we did not include the *Permuterm*<sup>®</sup> *Subject Index* and the *Corporate Index* that are part of the regular *SCI* service. The main reason for this is the large amount of translation work that would be required. Should this prove important in the future, it may be considered.

The new cumulation will cover more than 1,000,000 source journal items and more than 13,000,000 cited references. This information will be found exclusively in this printed version of the *SCI* cumulation. It will not be made available online until or unless the number of subscribers is sufficient to cover costs

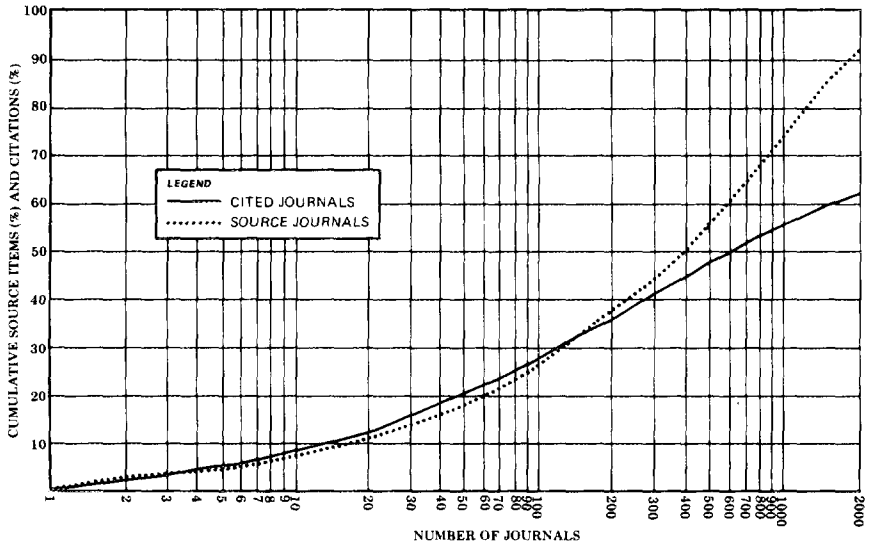
and to help finance the continued processing of pre-1955 material.

Those unfamiliar with certain scientometric principles might well wonder how completely a journal list of 500 can cover the literature of science. I won't repeat here in detail the origins of Bradford's law of scattering. In 1934,<sup>5</sup> Samuel Bradford studied the literature of electrical engineering and showed that it was divided into three zones. In the first zone, about ten journals would account for one third of the literature. To cover another third of the literature, you would need about 60 journals, and for the last third about 260.<sup>6</sup> A physical analogy of Bradford's law would be a comet. The nucleus represents the core journals of a field, while the tail represents the additional journals that sometimes publish material relevant to the subject. The tail becomes wider in proportion to the distance from the nucleus.

Bradford's law describes literature "scatter" in many different fields. But it says nothing about how the core journals of one field can overlap another field. This overlap is considerable. Not until the *SCI* data became available were we aware of the extent of this overlap. Our studies have shown that relatively few journals publish the majority of cited articles for all fields.<sup>7</sup> This observation formed the basis of Garfield's law of concentration.<sup>8</sup> I noted that quite often the "tail" of the literature of one discipline consisted of the "core" literature of other disciplines. When you combine all these tails and cores, there is a surprisingly small group of journals that account for the most significant literature for all of science. This is confirmed even more dramatically in our analysis of *Citation Classics*.<sup>9</sup>

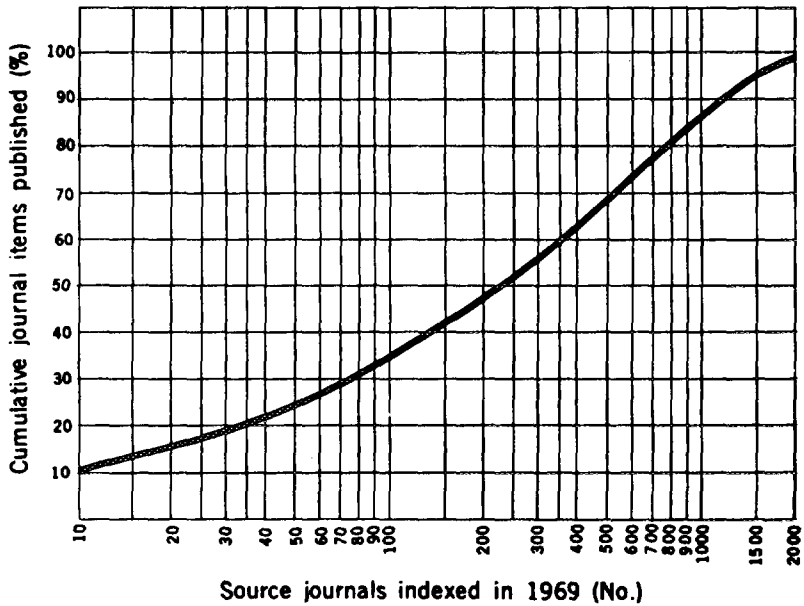
Figure 1 illustrates the law of concentration. The graph was prepared from data derived from the 1981 *SCI*. Ten journals account for about ten percent of what is published, and the same number account for ten percent of what is cited! Only seven percent, or 200 of the more than 3,000 journals *SCI* covered in

**Figure 1:** Distribution of published items and citations among science journals.



BASED ON 1981 SCI

**Figure 2:** Distribution of the number of published items among the approximately 2,200 journals covered by SCI® in 1969. A relatively small core of journals carried the majority of items published.



1981, published more than 35 percent of the articles indexed and cited. With just 500 journals in your collection, you can account for more than half of what is published and cited worldwide. From this point on, there is a gradually diminishing return by increasing the number of journals one covers. These tend to be smaller, more specialized journals.

Of course, the literature has grown tremendously in the past two decades. In 1969, for example, we found that just 500 journals out of the 2,200 then covered in *SCI* published nearly 70 percent of the articles indexed (see Figure 2). The same 500 journals accounted for nearly 70 percent of the citations. But even then, the top ten journals accounted for ten percent of what is published. I will not attempt to explain in detail some of the reasons why, at the 65 percent mark, adding more and more journals to our coverage does not significantly increase the number of journals that are well cited. Half the remaining citations do not involve journals. Much of the remainder are in fact multiauthored "books," or serials of one kind or another which are not taken into account in these studies. So we have good reason to believe that we have optimized the coverage of the 1955-1964 *SCI* cumulation by including 500 journals from that period.

As an important aid in the application of *SCI* to the detailed examination of

specific research fields, we are now considering the creation of a new index to research fronts for the period 1955-1964. Using the same basic techniques I have described elsewhere,<sup>2</sup> we could identify up to 2,000 research topics that were the most active during this decade. And for each research front, we would identify a group of the core papers. We would have more to say about these primordial papers in a follow-up discussion. At that time, we would provide a list of the most active clusters for the decade.

The 1955-1964 *SCI* cumulation will be published early in 1984 and will cost \$25,000. However, libraries can take advantage of a prepublication price of \$20,000. They may also trade in *SCI* annuals from 1961 to 1964, and receive a \$400 credit per annual toward the purchase price of the cumulation. If your institution is a recipient of an ISI grant, you can purchase the cumulation at your usual grant rates. All payments for the *SCI* cumulation may be made in installments. For more information about the 1955-1964 *SCI* cumulation, contact ISI Customer Services, 3501 Market Street, University City Science Center, Philadelphia, PA 19104.

\* \* \* \* \*

*My thanks to Thomas Di Julia for his help in the preparation of this essay.*

©1983 ISI

#### REFERENCES

1. **Garfield E.** *Citation indexing—its theory and application in science, technology, and humanities.* New York: Wiley, 1979. 274 p.
2. ...., ABCs of cluster mapping. Parts 1 & 2. Most active fields in the life and physical sciences in 1978. *Essays of an information scientist.* Philadelphia: ISI Press, 1981. Vol. 4. p. 634-49. (Reprinted from: *Current Contents* (40):5-12, 6 October 1980 and (41):5-12, 13 October 1980.)
3. ...., Introducing the *ISI Atlas of Science: Biochemistry and Molecular Biology, 1978/80.* *Current Contents* (42):5-13, 19 October 1981.
4. **Garfield E & Sher I H,** eds. *Genetics Citation Index.* Philadelphia: Institute for Scientific Information, 1963. 864 p.
5. **Bradford S C.** Sources of information on specific subjects. *Engineering* 137:85-6, 1934.
6. ...., *Documentation.* Washington, DC: Public Affairs Press, 1950. 156 p.
7. **Garfield E.** Citation analysis as a tool in journal evaluation. *Science* 178:471-9, 1972.
8. ...., The mystery of the transposed journal lists—wherein Bradford's law of scattering is generalized according to Garfield's law of concentration. *Essays of an information scientist.* Philadelphia: ISI Press, 1977. Vol. 1. p. 222-3. (Reprinted from: *Current Contents* (14):5-6, 4 August 1971.)
9. ...., *Citation Classics—four years of the human side of science.* *Current Contents* (22):5-16, 1 June 1981.