Essays of an Information Scientist, Vol:1, p.42, 1962-73
Current Contents, #9, August 26, 1969

"current comments"

Citation Indexes in Sociological and Historical Research

Click here to go to reprint :
"Citation Indexes in Sociological and Historical Research"
Volume 1, p.43

August 26, 1969

It seems incredible that more than fifteen years have passed since I first began my investigations on citation indexing. In a future editorial I should like to review the rocky history of this concept and how it finally emerged as a practical bibliographical tool used in most of the major scientific libraries of the world. However, it was in 1964 that ISI® made one of the most important financial decisions in its history - - launching the *Science Citation Index*® as a continuing service. It has taken over five years to turn that gamble into a reasonably profitable enterprise and with continued growth, *SCI*® should remain so for many years to come.

However, you will readily understand why, back in 1964, we were not prepared to add to that risk by widely publicizing the use of the *SCI* as a tool for evaluation of scientific performance. We had a number of bad experiences which proved, so it seemed, that if we pursued this application aggressively, the *SCI* might fail, not because it was a bibliographic failure - - on the contrary, because it was a success.

In its present stage of development, there is little danger that irate egocentrics can cause the *SCI* to be chucked from their libraries. Therefore, we should like to open Pandora's box. Let the scientific community decide for itself whether the *SCI* is a valid sociometric tool or not.

I should like to begin the process by reprinting one of my early papers on this subject in this issue. (See pp. 31). Reprints are available on request. Your comment is invited. In subsequent columns I plan to report on several other papers that have been published in this new field - - sometimes called the Science of Science. For those who are especially interested, several pertinent references are listed below. Reprints of asterisked items are available free.

*1) Garfield, E. & Sher, I.H., "New Tools for Improving and Evaluating the Effectiveness of Research", Yovits, M.C., ED., and others, *Research Program Effectiveness*, Proceedings of the Conference Sponsored by the Office of Naval Research, Washington, D.C., July 27-29, 1965 (New York: Gordon and Breach, 1966), pp. 135-146

*2) Garfield, E. & Sher, I.H., "New Factors in the Evaluation of Scientific Literature Through Citation Indexing" *American Documentation*, Vol. 14(3), pp. 195-201 (1963)

*3) Garfield, E., "Citation Indexes in Sociological and Historical Research," *American Documentation*, Vol. 14 (4), pp. 289-291 (1963)

4) Garfield, E., Sher, I.H., Torpie, R.J., *"The Use of Citation Data in Writing the History of Science"*, (Philadelphia: Institute for Scientific Information, 1964).

*5) Margolis, J., "Citation Indexing & Evaluation of Scientific Papers" *Science*, Vol. 155, pp. 1213-9 (1967).

Essays of an Information Scientist, Vol:1, p.43-46, 1962-73 Current Contents, #9, August 26, 1969
Reprinted from : American Documentation, (14) p.289-91, 1963
Reprinted from *American Documentation* 14(4):289-291, October 1963.

Citation Indexes in Sociological and Historical Research

by
Eugene Garfield
Institute for Scientific Information
Philadelphia, Pa. 19106

The use of citation data in the computer construction of historical network diagrams is proposed. A new 'critical path method' of evaluating the impact of individual scientific discoveries is described as well as a 'critical impact' factor which measures the arborization of specific papers or ideas through the subsequent literature. The danger of indiscriminate, unqualified use of quantitative citation data in sociological evaluations is emphasized. However, citation indexes do make thorough qualitative evaluations possible by providing a practical tool for locating subsequent criticisms of particular papers by specified authors.

In 1955 I proposed the compilation of comprehensive citation indexes primarily as an effective means of disseminating and/or retrieving scientific literature.[1] Recently, a project was begun whose specific purpose is to determine the value of citation indexing for the widely dispersed literature of genetics.[2] Preliminary sample citation indexes compiled from 326,000 citations appearing in 1960 journals were prepared. One was an author citation index in which approximately 230,000 different papers were cited an average of 1.4 times, the range extending from 1 to 304. A paper by Lowry on protein determination[3] was most frequently cited. Some preliminary detailed statistics drawn from this file were recently reported.[4] A journal citation index for the same 326,000 citations was also compiled and includes statistics on the frequency of citation for each journal cited.

A file of over 1.4 million citations from the 1961 literature has now been compiled in which an estimated 750,000 different papers have been cited. Every citation appearing in 613 journals is included. Major categories covered include the interdisciplinary journals (*Science, Nature, Doklady, Proc. Natl. Acad. Sci.*), instrumentation, physical chemistry and physics, clinical and experimental medicine, chemistry, genetics, and life sciences such as microbiology.

In the summer of 1963, a computer-generated genetics citation index will be extracted from this file and published

This work has been supported in part by the National Science Foundation, Office of Science Information Service, under contract No. C-201, and by the National Institutes of Health under Grant No. RG 8050.

for evaluation by 1,000 geneticists. The selection will be primarily those papers which have cited the work of a list of geneticists, papers which have appeared in "genetics" journals, and papers recommended by an Advisory Committee of geneticists.[5]

The display of the 1960 sample citation index at several scientific meetings and considerable correspondence with scientists have generated a high degree of interest in citation indexing for retrieval and dissemination purposes. This will be reported on in detail later. However, interest in using citation indexes for retrieval and dissemination is equaled, if not exceeded, by the interest in its use for sociological and historical research. Indeed, Newell stated: "Citation Indexing will generate a spate of empirical work on the sociology of science.... It is rather easy to predict, I think, that the publication and wide availability of an extensive citation index will have strong social consequences along the line of becoming a controlling variable for the advancement and employment of scientific personnel.... It makes little difference whether one likes this or not."[6] Merton said: "I am persuaded that your materials should be a rich source for the sociologist of science. As it happens, I am now in the midst of working on a problem in this field which needs precisely the kind of evidence you are putting together in your Citation Index."[7]

One purpose of this communication is to record my forewarning concerning the possible promiscuous and careless use of quantitative citation data for sociological evaluations, including personnel and fellowship selection. In particular, I wish to disassociate myself from such abuse of citation data recently imputed to my by Swanson.[8] He erroneously stated that in my 1955 paper in *Science*[1] I claimed one could measure the *importance* of a paper by citation counting. Citation counting is an old technique[9] and has been criticized for many reasons by Brodman,[10] Raisig,[11] and others. *Impact* is not the same as *importance* or *significance*. There is no specific correlation between the number of papers published by an individual and the quality or importance of his work, though Price[12] has indicated that scientists who produce work of high quality *usually* have a high publication rate. We can confirm this and add the observation that their papers *usually* are also cited more frequently than the average.

Westbrook[13] has justifiably used citations as a means of determining the sources of activity in particular research fields, and citation indexing clearly and demonstrably has proven to be an excellent means of "defining" a field. We hope to improve citation technology and have worked out many corrective factors, but the data presently available are inconclusive.

Citation indexes can be used to *facilitate* personnel and fellowship evaluation simply because they provide more convenient access to the literature. Citation indexes synthesize a consensus of scientific opinion needed in a careful appraisal of research, whether for editorial refereeing, making awards, or selecting personnel. It is preposterous to conclude blindly that the most cited author deserves a Nobel prize. On this basis, Lysenko and others might have been

judged the greatest scientists of the last decade. Such quantitative data can have *significance* for the historian who can carefully evaluate all the data available. Surely, the history of science must record the controversial as well as the noncontroversial figure. However, the mere ranking by numbers of citations or the numbers of papers published is no way to arrrive at objective criteria of importance.

Frequent use of citation files during the past 8 years has clearly indicated the enormous potential value of large-scale citation index files in facilitating historical research. This potential value has been confirmed in discussions with Bernal,[14] Shryock,[15] Leake,[16] and Price,[17] among others.[6,7] My specific interest in citation data for historical research has focused on the role of the computer in constructing historical "maps." The development of many sophisticated computer programs employing the critical path methodology for the Project Evaluation Review Technique (PERT) [18,19] stimulated the further pursuit of the now proposed technique in which the computer actually draws the topological network diagrams which show the chronological and derivational relationships between scientific papers and therefore scientific discoveries. Bernal prepared a similar small-scale map in tracing the antecedents and consequences of Pasteur's discovery of molecular asymmetry.[20] Allen prepared a similar network diagram for a bibliography on "Behavior of biological stains with special reference to staining of nucleic acids" which shows the historical relationships between each of the 15 papers listed.[21] However, the proposed critical path method does not stop at the graphical presentation of the citational interrelationships—in itself an otherwise tremendous chore for the historian.

One can also program the computer to "erase" one or more nodes and its connections in the network. By making the assumption that any papers which derive from these nodes should also be erased, one can then determine whether an alternate "critical path" to a subsequent discovery does exist. Thus, one can legitimately ask whether the work of Whitnah on the physical properties of milk[22] or Elworthy's work on lecithin micelles[23] would have been possible if Einstein had not published his paper on measuring dimensions of molecules![24] Further, by counting the total number of "terminal" nodes erased, one *may* possibly obtain a useful quantitative measure of historical impact. A terminal node is a member of a particular set of papers, as, e.g., all 1962 papers. The critical impact of Einstein's work would be the percentage of 1962 papers which, one assumes, would not have been written if they could not have cited *either* Einstein's papers *or* papers which have cited Einstein. In making such calculations, one has to also consider the papers cited by Einstein. It is a common practice for authors to cite, and sometimes exclusively, the papers cited in the original work which inspired the ideas discussed. Thus, to trace the impact of my 1955 paper, one has to also consider citations to my deceased colleague and friend, W.C. Adair, who suggested citation indexing to me.[25] It is significant that, for similar reasons, when using citation indexes for literature searching,

the search strategy frequently requires examination of references not only to a specific paper. It may frequently be necessary, if not preferable, to examine references to the papers in the bibliography of the target paper. This technique is called *cycling*.[5] If S are the sources which cite one or more members of the bibliography B in target reference R, then S[B(R)] is the pertinent list of papers. This method is quite different from searching by *bibliographic coupling* which can, of course, be done with a printed citation index. In this case one retrieves S(Rn) where n is the degree of coupling or the number of different papers cited in common by two or more sources.[26]

1. **Garfield E.** Citation indexes for science. **Science** 122:108, 1955.
2. Institute for Scientific Information®. **J. Heredity** 52:182, 1961.
3. **Lowry O H et al.** Protein measurement with the Folin phenol reagent. **J. Biol. Chem.** 193:265, 1951.
4. **Garfield E & Sher I H.** New factors in the evaluation of scientific literature through citation indexing. **Amer. Documentation** 14:195-201, 1963.
5. **Allen G, Cavalli-Sforza L, Lederberg J, LeFevre G, Melnick J & Spiegelman.** *Research and evaluation program on citation indexing.* (Philadelphia: Institute for Scientific Information, 1962), 5 pp.
6. **Newell A.** Private communication, 27 February 1962.
7. **Merton R K.** Private communication, 19 April 1962.
8. **Swanson D R.** Library goals and the role of automation. Paper presented at the 16 June 1962 meeting of the Association of Research Libraries, Miami Beach, Florida. (Note: Dr. Swanson eliminated his reference to citation indexing in the subsequently published version of this talk; cf. **Special Libraries** 53:466-71, 1962.
9. **Gross P L K & Gross E M.** College libraries and chemical education. **Science** 66:385, 1927.
10. **Brodman E.** Choosing physiology journals. **Med. Libr. Assoc. Bull.** 32:479, 1960.
11. **Raisig L M.** Mathematical evaluation of the scientific serial. **Science** 131:1417, 1960.
12. **Price D J D.** *Little science, big science.* (New York: Columbia Univ. Press, 1963), p. 41.
13. **Westbrook J H.** Identifying significant research. **Science** 132:1229, 1960.
14. **Bernal J D.** Private communication, 22 March 1962.
15. **Shryock R.** Private conversation, 19 September 1962.
16. **Leake C D.** Private communication, 29 August 1962.
17. **Price D J D.** Private communication, 15 March 1962.
18. **Lasser D J.** Topological ordering of a list of randomly numbered elements of a network. **Commun. ACM** 4:167, 1961.
19. **Kahn A B.** Topological sorting of large networks. **Commun. ACM** 5:558, 1962.
20. **Bernal J D.** *Science and industry in the nineteenth century.* (London: Routledge Kegan Paul, 1953), p. 230.
21. **Allen G.** Private communication, 2 June 1960.
22. **Whitnah C H & Rutz W D.** Some physical properties of milk. 6. The voluminosity of caseinate complex in milk, and reconstituted sediments. **J. Dairy Science** 42:227, 1959.
23. **Elworthy P H.** The structure of lecithin micelles in benzene solution. **J. Chem. Soc.** p. 1951, 1959.
24. **Einstein A.** Eine neue Bestimmung der Molekuldimensionen. **Ann. Physik** 19:289, 1906.
25. **Adair W C.** Citation indexes for scientific literature? **Amer. Documentation** 6:31, 1955.
26. **Kessler M M.** Bibliographic coupling between scientific papers. **Amer. Documentation** 14:10, 1963.