Citation Indexing for Studying Science

by
E. GARFIELD
Institute for Scientific Information,
Philadelphia, Pennsylvania 19106

By revealing who has really influenced the course of science the *Science Citation Index* seems to be a valuable sociometric tool for historians and sociologists.

By writing the *Double Helix*[1], Watson laid to rest the absurd notion that scientists have less desire for reputational immortality than other humans. That this belief could have existed at all is only a sign that the efforts of scientists to achieve fame and fortune are necessarily less obvious than those of athletes and politicians. We are thus confronted by a situation wherein those scientists who deserve (and want) recognition cannot always be easily identified, even by their peers. It seems likely, then, that social scientists—whose role is to tell it like it is—will begin to play a larger part in identifying those scientists who have had or will have a major impact on their fields. Indeed, a new breed of sociometrist is developing, called the scientist of science[2], that is concerned chiefly with the historical, sociological, economic and behavioural study of science and scientists.

Unfortunately, the measurement of science will not become more precise, even though there is a specialized group doing the measuring, unless more effective measuring techniques are developed and used. Most evaluation procedures available to sociometrists are not only slow and costly, they are also tedious. Such practices as counting the number of papers published have been used because truly objective methods were not available. The exponential growth of scientific research and the increasing number of scientists only make matters worse. It is in meeting this need for an effective, efficient, and unobtrusive[3] sociometric tool that citation indexing may find its most important application.

A citation index is an ordered list of cited articles, each accompanied by a list of citing articles. The citing article is identified as a source, the cited article as a reference[4]. The *Science Citation Index* (SCI), published by the Institute for Scientific Information, is the only regularly issued citation index in science. It is prepared by computer and provides an index to the contents of every issue published during a calendar year of more than 2,000 selected journals. Journals covered by the index are chosen by advisory boards of experts in each of the topics represented and by large scale citation analyses.

The entry for a cited article (reference) contains the author's name and initials, the cited reference year, and the publication name, volume, and page number. Under the name of each cited author appears the source article citing this work. This line is arranged by citing author's name, publication, type of source item (article, abstract, editorial and so on), citing year, volume, and page. The searcher starts with a reference or an author he has identified through a footnote, book, encyclopaedia, or conventional word or subject index. He then turns to the *Citation Index* section of the SCI and searches for that particular author's name. When he has located the name, he checks to see which of several possible references fits the particular one he is interested in. He then looks to see who has currently cited this particular work. After noting the bibliographic citations of the authors who are citing the work with which he started, the searcher then turns to the *Source Index* of the SCI to obtain the complete bibliographic data for the works which he has found.

After finding several source articles, the searcher can use the bibliographies of one or several of these as other entries into the citation index; this process is called "cycling". Since authors frequently write more than one closely related paper, additional articles by the author of the starting reference can also be used as entry points to the index.

Basically, then, the SCI does two things[5]. First, it tells what has been published. Each annual cumulation cites between 25 and 50 per cent of the 5 to 10 million papers and books estimated to have been published during the entire history of science. Second, because a citation indicates a relationship between a part or the whole of a cited paper and a part or the whole of the citing paper, the SCI tells how each brick in the edifice of science is linked to all the others[6]. Because it performs these two fundamental functions so well, important applications for the SCI have been found in three major areas: library and information science, history of science, and the sociology of science.

The SCI was originally designed to be a retrieval tool for use in library and information science work[7]. It has served this purpose very well. The unique retrieval effectiveness of the SCI has already been reported by several investigators[8]. The worldwide adoption of SCI in its short history confirms its ability to augment traditional indexing methods.

Uses in Historical Research

Besides retrieval, other uses for the SCI in library and information science are emerging. Because well over 20 million bibliographic citations have been extracted from more than 1,500,000 source documents, the SCI data base can be utilized to provide definitive studies of journal-to-journal relationships. A recent study by Martyn[9] illustrated how the SCI data base could be used

to rank British scientific journals and pick out the effective "hard core" of literature. Soon, the Institute for Scientific Information will publish a statistical compilation which will show how often each of 2,000 journals cite one another. This *Source Journal Citation Index* will be complemented by the *Reference Journal Citation Index*, which will show how often each of these 2,000 journals cites any of more than 25,000 other journals.

The suggestion for using citation indexing for historical research came as early as 1955[10]. Dr Gordon Allen gave great impetus to this idea when he constructed a bibliographic citation network diagram in 1960 (personal communication from G. Allen). In 1964 the practical methodology was developed to permit the use of citation indexing in sociological and historical research to identify key events, their chronology, their interrelationships, and their relative importance[11].

Fig. 1 shows the application of SCI data to create a graphic aid to the study of the history of science. By examining the interconnecting links of scientific events shown in the citation network, it is possible to observe historical and sociological processes at work. It is also easy to identify the nodal publications in the citation network, that is, those that are cited most by others, those that have had the most impact. From Fig. 1 it would be quite reasonable to conclude that whoever published paper number 2 had considerable impact on research involving nucleic acid staining. It is at this point that the SCI begins to serve as an objective sociometric tool; it begins to show who has truly influenced the course of science.

In addition to identifying individuals whose work has had impact on a branch of science, carefully constructed citation networks can help disprove certain prevailing scientific myths. For example, it is commonly believed that Gregor Mendel's breakthrough paper on genetics was ignored by the scientific community from the time it was presented in 1865 until it was "rediscovered" in 1900. The citation network in Fig. 2 shows, however, that not only was Mendel's work not ignored, but that it was actually cited by at least four different people before 1900. Mendel's work was even cited in an article on "hybridism" in the ninth edition of the *Encyclopaedia Britannica*. One could hardly call that being ignored.

Citation networks can also bring into focus anomalies in the history of scientific development. In Fig. 2, for example, why did Darwin's 1876 paper cite Hoffman but not Mendel? Certainly this is unusual, since Hoffman's paper cites Mendel five times. Inconsistencies like these are clearly identified in citation networks and give impetus and assistance to all types of important historical research.

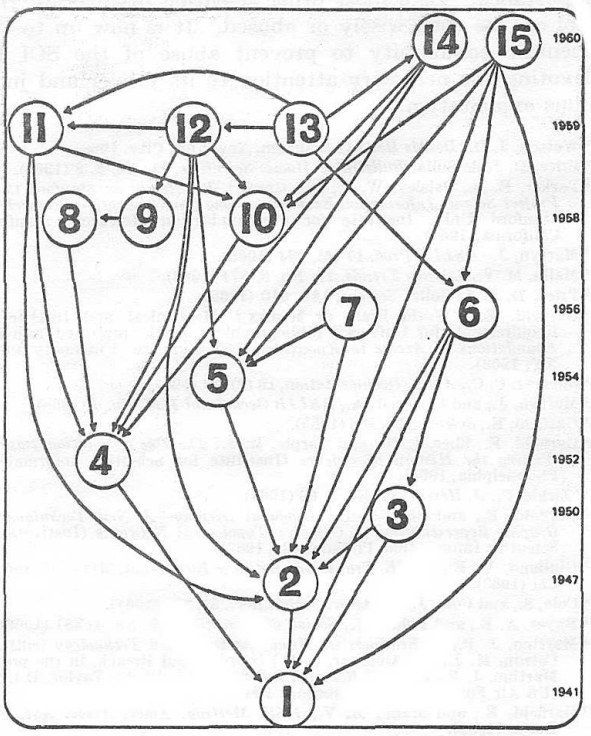

Fig. 1. Citation network of fifteen articles on nucleic acids.

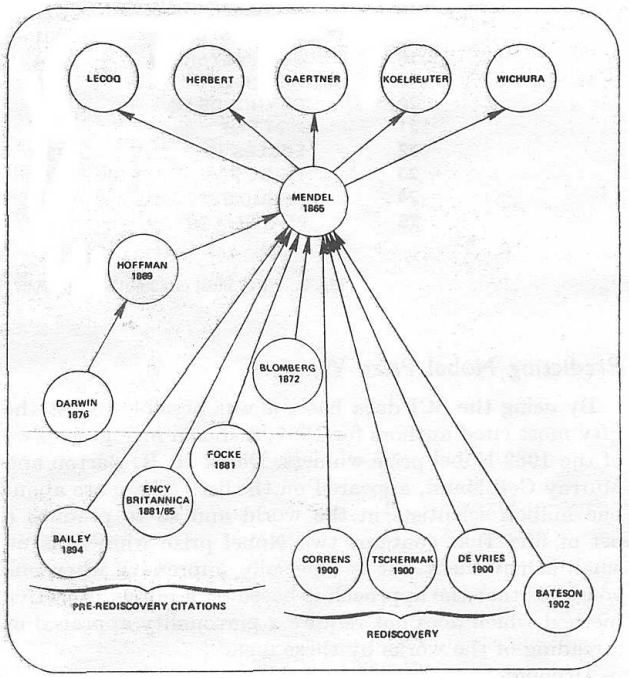

Fig. 2. Citation network showing citations to Mendel before alleged "rediscovery". This was discussed by Zirkle in ref. 11*a*.

The citation networks shown were produced manually, but further work[12] indicates that such diagrams can be assembled automatically using large computer memories and programs for iterative display of appropriate data. This means that in the near future a historian or sociometrist will be able to sit before a computer console and specify some starting point—a person, a word, a citation, or a place[13]. He will then ask the computer to display a list of pertinent papers. The computer will respond by drawing or displaying a historical road map which will show not merely a list of papers and books, but also a graphical approximation of the history of that subject.

It was a logical step to progress from using the SCI as a sociometric tool in historical contexts to using it to measure current scientific performance[14]. Bayer[15], Martino[16], and others have already reported that valid correlations can be obtained between individual performance and citation counts. Perhaps the most dramatic indication of the sociometric power of the SCI was the forecast made in 1968 of those who would win Nobel prizes in 1969[17].

RANK	AUTHOR	TOTAL TIMES CITED	RANK	AUTHOR	TOTAL TIMES CITED
1	LOWRY OH	2921	26	ELIEL EL	721
2	CHANCE B	1374	27	STREITWIESER A	717
3	LANDAU LD	1174	28	MULLIKEN RS	712
4	BROWN HC	1150	29	JACOB F	71
5	PAULING L	1063	30	BORN M	71
6	GELLMANN M	942	31	BRACHET J	706
7	COTTON FA	940	32	WINSTEIN S	70
8	POPLE JA	933	33	ALBERT A	68
9	BELLAMY LJ	906	34	LUFT JH	67
10	SNEDECOR GW	904	35	DEDUVE C	67
11	BOYER PD	893	36	VONEULER US	66
12	BAKER BR	876	37	FIESER LF	666
13	KOLTHOFF IM	853	38	HUISGEN R	66
14	HERZBERG G	842	39	NOVIKOFF AB	655
15	FISCHER F	826	40	GOODWIN TW	643
16	SEITZ F	822	41	BARTON DHR	632
17	DJERASSI C	801	42	FISHER RA	631
18	BERGMEYER HU	754	43	BATES DR	627
19	WEBER G	750	44	FLORY PJ	626
20	REYNOLDS ES	748	45	STAHL E	626
21	MOTT NF	741	46	DEWAR MJS	615
22	ECCLES JC	737	47	GILMAN H	618
23	FEIGL F	729	48	FOLCH J	618
24	FREUD S	727	49	DISCHE Z	614
25	PEARSE AGE	726	50	GLICK D	60

Nobel Prize in Physics 1969

Nobel Prize in Chemistry 1969

Fig. 3. Fifty most cited authors for 1967, ranked according to total number of times cited.

Predicting Nobel Prize Winners

By using the SCI data base, it was possible to list the fifty most cited authors for 1967 as shown in Fig. 3. Two of the 1969 Nobel prize winners, Derek H. R. Barton and Murray Gell-Mann, appeared on the list. There are about one million scientists in the world and so to produce a list of fifty that contains two Nobel prize winners is no small achievement. It is especially impressive when one considers that the approach is based on a purely objective method which does not require a personality appraisal or a reading of the works by these men.

Although forecasting Nobel prize winners is an interesting exercise, the ability of the SCI to identify those individuals who will make a major impact on science has more practical social and economic consequences. Research administrators in academic, industrial, and government organizations have frequently indicated the need for a tool for identifying such people. Increasingly scarce intellectual and financial resources for supporting research could be managed more efficiently with such an identification tool. Creative people could be identified much earlier in their careers so that they could benefit from special training. Prizes, grants, fellowships and other forms of recognition could be awarded without the wasteful in-fighting and manoeuvring among scientists described by Watson[1].

Another problem facing research administrators is how to determine the directions research should take in the future. The recent summary of the difficulties involved in selecting lunar experiments for the Apollo program[18] is a good current example of this type of dilemma.

In this kind of situation, imaginative use of the SCI data base might contribute to a solution. In the near future ISI will publish what should prove to be a valuable forecasting tool. This will be a regularly published list of the 20,000 papers which are cited most in a given year. Proper analysis of this information could be a giant step forward in identifying "where the action is" (or should be) in the area of scientific research.

When the *Science Citation Index* was first proposed, its major objective was to break the so-called subject index barrier[19]. Out of this bibliographic experiment has evolved a historiographic and sociometric tool of major importance. Like most other scientific discoveries, this tool can be used wisely or abused. It is now up to the scientific community to prevent abuse of the SCI by devoting the necessary attention to its proper and judicious exploitation.

[1] Watson, J. D., *Double Helix* (Athaneum, New York City, 1968).
[2] Price, D. J. de Solla, *Bulletin of Atomic Scientists*, 21, No. 8, 2 (1965).
[3] Parker, E. B., Paisley, W. J., and Garrett, R., *Report on Stanford Univ. Project Science Information Exchange among Communication Researchers*, Stanford Univ. Institute for Communication Research, Stanford, California, 1967).
[4] Martyn, J., *ASLIB Proc.* 17 (6), 184 (1965).
[5] Malin, M. V., *Library Trends*, 16, No. 3, 374 (1968).
[6] Price, D. J. de Solla, *Science*, **149**, 510 (1965).
[7] Garfield, E., "World Brain or Memex? Mechanical and Intellectual Requirements for Universal Bibliographic Control," reprinted from *The Foundations of Access to Knowledge*, 169 (Syracuse University Press, NY, 1968).
[8] Spencer, C. C., *Amer. Documentation*, **18** (2), 87 (1967).
[9] Martyn, J., and Gilchrist, A., *ASLIB Occasional Pub. No. 1* (1968).
[10] Garfield, E., *Science*, **122**, 108 (1955).
[11] Garfield, E., Sher, I. H., and Torpie, R. J., *The Use of Citation Data in Writing the History of Science* (Institute for Scientific Information, Philadelphia, 1964).
[11a] Zirkle, C., *J. Hered.*, **55**, No. 2, 65 (1964).
[12] Garfield, E., and Sher, I. H., *Diagonal Display—A New Technique for Graphic Representation of Complex Topological Networks* (Institute for Scientific Information, Philadelphia, 1967).
[13] Giuliano, V. E., *IEEE Trans. on Military Electronics*, MIL-7 (2 and 3), 221 (1963).
[14] Cole, S., and Cole, J. R., *Amer. Sociol. Rev.*, **32**, 377 (1967).
[15] Bayer, A. E., and Folger, J., *Sociol. of Education*, **39**, No. 4, 381 (1966).
[16] Martino, J. P., in *Readings in Managing Organized Technology* (edit. by Cetron, M. J., and Goldhar, J. D.) (Gordon and Breach, in the press); Martino, J. P., *AFOSR Research AD* 659 366 (edit. by Taylor, D.), 226 (US Air Force Office of Aerospace Research, Arlington, 1967).
[17] Garfield, E., and Malin, M. V., *135th Meeting, Amer. Assoc. Adv. Sci.*, Dallas (1968).
[18] *Nature*, **224**, 529 (1969).
[19] Garfield, E., *J. Patent Office Soc.*, 39, No. 8, 583 (1957).