

Citation Indexing for Studying Science

Eugene Garfield

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 behavioral 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 unobtrusive³ 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, 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 Martyn⁹ 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.^{10a} 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

Figure 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 Figure 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 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 Figure 2 shows, however, that not only was Mendel's work not ignored, but that it was actually

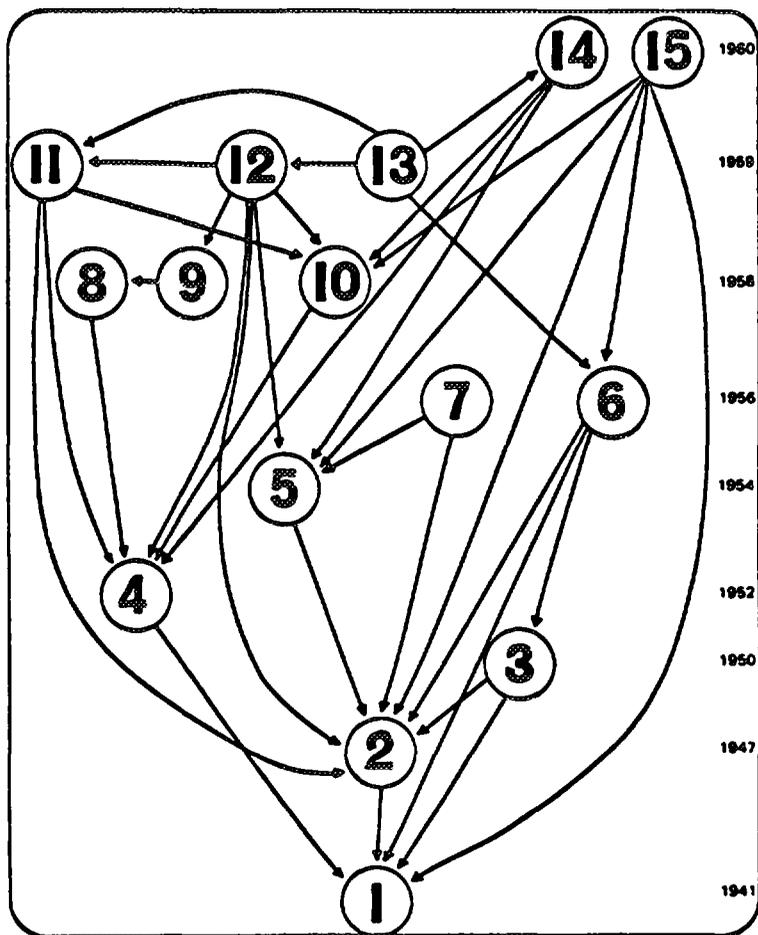


Figure 1. Citation network of fifteen articles on nucleic acids.

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 Figure 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.

The citation networks shown were produced manually, but further work¹² 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

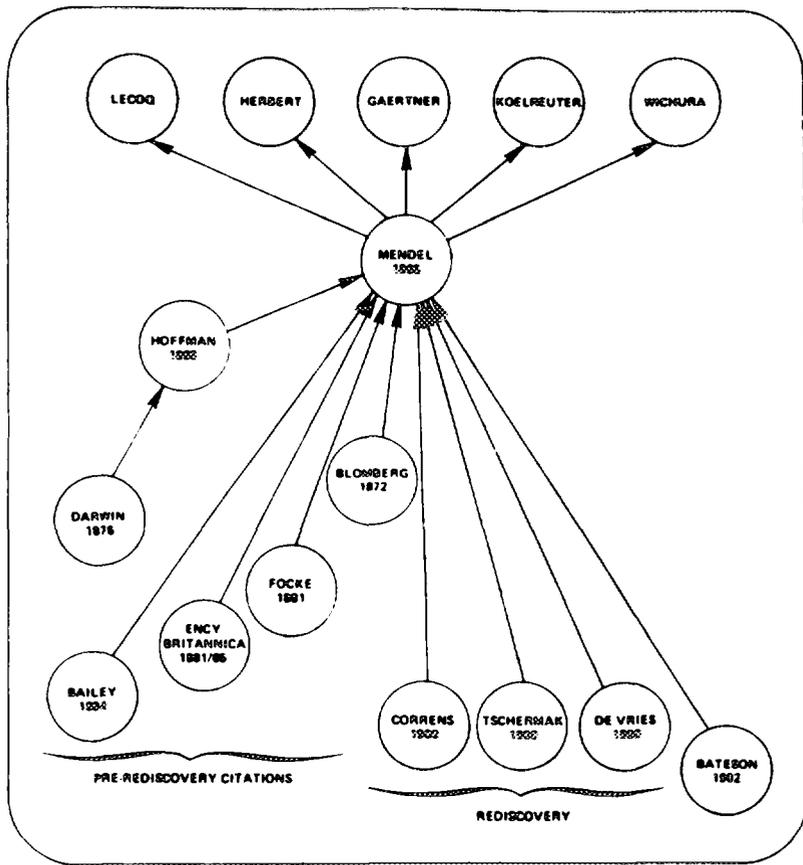


Figure 2. Citation network showing citations to Mendel before alleged "rediscovery". This was discussed by Zirkle in ref. 11a.

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

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 Figure 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

Figure 3. Fifty most cited authors for 1967, ranked according to total number of times cited. *Nobel Prize in Physics, 1969. †Nobel Prize in Chemistry, 1969. Authors in boldface type have won Nobel Prizes in later years.

Times			Times			Times		
Rank	Cited	Name	Rank	Cited	Name	Rank	Cited	Name
1	2921	Lowry OH	18	754	Bergmeyer HU	35	673	Deduve C
2	1374	Chance B	19	750	Weber G	36	668	von Euler US
3	1174	Landau LD	20	748	Reynolds ES	37	666	Fieser LF
4	1150	Brown HC	21	741	Mott NF	38	661	Huisgen R
5	1063	Pauling L	22	737	Eccles JC	39	655	Novikoff AB
6	942	Gell-Mann M *	23	729	Feigl F	40	643	Goodwin TW
7	940	Cotton FA	24	727	Freud S	41	632	Barton DHR †
8	933	Pople JA	25	726	Pearse AGE	42	631	Fisher RA
9	906	Bellamy LJ	26	721	Eliet EL	43	627	Bates DR
10	904	Snedecor GW	27	717	Streitwieser A	44	626	Flory PJ
11	893	Boyer PD	28	712	Mulliken RS	45	626	Stahl E
12	876	Baker BR	29	711	Jacob F	46	619	Dewar MJS
13	863	Kolthoff IM	30	710	Born M	47	618	Gilman H
14	842	Herzberg G	31	706	Brachet J	48	618	Folch J
15	826	Fischer F	32	702	Winstein S	49	614	Dische Z
16	822	Seitz F	33	687	Albert A	50	609	Glick D
17	801	Djerassi C	34	674	Luft JH			

based on a purely objective method which does not require a personality appraisal or a reading of the works by those 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 program¹⁸ 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.

References

1. **Watson J D.** *Double helix* (New York: Athenaeum, 1968).
2. **Price D J D.** The science of science. *Bull. Atomic Scientists* 21(8):2-8, 1965.
3. **Parker E B, Paisley W J & Garrett R.** *Report on Stanford University research project 'Science information exchange among communication researchers.'* (Stanford: Stanford University Institute for Communication Research, 1967).
4. **Martyn J.** An examination of citation indexes. *ASLIB Proceedings* 17(6):184-96, 1965.
5. **Malin M V.** The *Science Citation Index*, a new concept in indexing. *Library Trends* 16(3):374-87, 1968.
6. **Price D J D.** Networks of scientific papers. *Science* 149:510-15, 1965.
7. **Garfield E.** World brain or Memex? Mechanical and intellectual requirements for universal bibliographic control. In: *Foundations of Access to Knowledge* ed. **E B Montgomery** (Syracuse, New York: Syracuse University Press, 1968), p. 169-96.
8. **Spencer C C.** Subject searching with the *Science Citation Index*; preparation of a drug bibliography using *Chemical Abstracts*, *Index Medicus*, and *Science Citation Index* 1961 and 1964. *American Documentation* 18(2):87-96, 1967.
9. **Martyn J & Gilchrist A.** *An evaluation of British scientific journals.* ASLIB Occasional Publication Number 1. (London: ASLIB, 1968).
10. **Garfield E.** Citation indexes for science. *Science* 122:108-11, 1955.
- 10a. **Allen G.** Personal communication.
11. **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).
- 11a. **Zirkle C.** Some oddities in the delayed discovery of Mendelism. *J. Heredity* 55:65-72, 1964.
12. **Garfield E & Sher I H.** *Diagonal display; a new technique for graphic representation of complex topological networks.* (Philadelphia: Institute for Scientific Information®, 1967).
13. **Giuliano V E.** Analog networks for word association. *IEEE Trans. Military Electronics MIL-7*:(2 & 3): 221, 1963.
14. **Cole S & Cole J R.** Scientific output and recognition; a study in the operation of the reward system in science. *Amer. Sociol. Rev.* 32:377-90, 1967.
15. **Bayer A E & Folger J.** Some correlates of a citation measure of productivity in science. *Sociology of Education* 39:381-90, 1966.
16. **Martino J P.** Citation indexing for research and development management. In: *Readings in managing organized technology* ed. **M J Cetron & J D Goldhar** (New York: Gordon & Breach, in press). Also: **Martino J P.** Research evaluation through citation indexing. AFOSR Research AD 656 366, ed. **D. Taylor** (Arlington, VA.: USAF Office of Aerospace Research, 1967).
17. **Garfield E & Malin M V.** Can Nobel Prize winners be predicted? Paper presented at the 135th Annual Meeting of the American Association for the Advancement of Science, Dallas, Texas, 1968.
18. **Anonymous.** No hats in the air for Apollo. *Nature* 224:529, 1969.
19. **Garfield E.** Breaking the subject index barrier; a citation index for chemical patents. *J. Patent Office Soc.* 39:583-95, 1957.