"Current Comments"

Citation Indexing and the Sociology of Science

Click for : "Citation Indexing, historio-bibliography, and the sociology of science", p158 Essays, Vol:1

April 14, 1971

The delay in publication of scientific papers is a constant source of frustration for their authors. Perhaps no segment of the literature is subjected to greater publication delays than that which eventually appears in the bound volumes emanating from international meetings and symposia. In May 1969, I presented a paper[1] which brought together much of my theoretical and practical work on the subject of indexing. During the twenty months it took to publish that work I was not scooped, as so often happens these days, but a number of developments did take place which made it obsolete without an appropriate supplement. I tried to rectify the situation by publishing a short paper in *Nature*[2] which has been reprinted in *Current Currents*® [3]. Indeed, the subject has been anonymously editorialized (a regrettable practice) in *Nature*[4] and amplified by Cawkell[5]. I in turn have responded to the challenge with a letter[6] whose publication has been delayed by the British postal strike.

Of even greater significance, insofar as the topic concerns science policy studies or the sociology of science, there have appeared in sociology journals not usually seen by most *CC*® readers a series of studies which provides almost incontrovertible support for the claim that citation analysis can be correlated quite well with other more subjective methods of analysis. Perhaps the most striking study to follow up the work of Cole[7] is that of Hagstrom[8], who has obtained an amazingly high correlation with Cartter[9] in evaluating graduate academic departments. These and several other relevant papers[10-15] are listed below and will be discussed in the future.

Following this editorial, the paper I originally read in Amsterdam is reprinted in its entirety. Quite frankly, this is the most economical method of getting it into the hands of those who have expressed interest in the topic. As is our custom, reprints are available.

1. Garfield, E. "Citation Indexing, Historio-Bibliography, and the Sociology of Science." in *Proceedings of the Third International Congress of Medical Librarianship, Amsterdam, 5-9 May, 1969*, ed. by K.E. Davis & W.D. Sweeney (Excerpta Medica, Amsterdam, 1970) pp. 187-204.
2. Garfield, E. Citation indexing for studying science. *Nature* 227:669-671, 1970.
3. *Current Contents/Life Sciences* 13(46):45-51, November 18, 1970.
4. Anonymous. More games with numbers. [An editorial in] *Nature* 228(5273):698-699, 1970.
5. Cawkell, A.E. *Science Citation Index*. [A letter to the editor of] *Nature* 228(5273): 789-790, 1970.

6. Garfield, E. Where the action was, is and will be. *Nature*, in press.

7. Cole, S. & Cole, J.R. Visibility and the structural bases of awareness of scientific research. *American Sociological Review* 33(3):397, 1968.

8. Hagstrom, W.O. "Inputs, Outputs, and the Prestige of American University Science Departments." Paper presented at the Annual Meeting of the American Association for the Advancement of Science, Chicago, December 28, 1970.

9. Cartter, A.M. *An Assessment of Quality in Graduate Education.* (American Council on Education, Washington, D.C., 1966).

10. Cole, S. & Cole, J.R. Scientific output and recognition; a study in the operation of the reward system in science. *American Sociological Review* 32(3):377-390, 1967.

11. Bayer, A.E. & Folger, J. Some correlates of a citation measure of productivity in science. *Sociology of Education* 39(4):382-390, 1966

12. MacRae, D. Jr. Growth and decay curves in scientific citations. *American Sociological Review* 34(5):631-635, 1969.

13. Parker, E.B., Paisley, W.J. & Garrett, R. "Bibliographic Citations as Unobtrusive Measures of Scientific Communication." (Stanford University Institute for Communications Research, Stanford, 1967, 125 pp.)

14. Price, D.J. deS. Is technology historically independent of science? A study in statistical historiography. *Technol. Culture* 6(4):553-568, 1965.

15. Whitley, R.D. Communication nets in science; status and citation patterns in animal physiology, *Sociological Review* 17(2):219-233, 1969.

Essays of an Information Scientist, Vol:1, p.158-174, 1962-73 Current Contents, #6, April 14, 1971 Reprinted from: Proceedings of the third international congress of medical librarianship (Amsterdam: Excerpta Medica, p.187-204, 1970)

citation indexing, historio-bibliography, and the sociology of science

by Eugene Garfield, Ph.D., President
Institute for Scientific Information

It is indeed an honor to have been asked by the Scientific Committee to replace my friend and colleague, Professor Derek de Solla Price, as the speaker on this occasion. I gladly accepted the challenge, but I cannot provide his unique blend of wit, humor, and scholarship. Both Professor Price and Professor Robert K. Merton serve on the Advisory Board of the *Science Citation Index* as representatives of the 'Scientists of Science' — the name for a new breed of sociometrist concerned with the historical, sociological, economic, and behavioral study of science and scientists.

In contrast to Price who has 'turned' from history to bibliography, or Merton who has similarly 'turned' from sociology to find gold in the hills of bibliotopia, I am the bibliographer turned historiographer and sociometrist. I, therefore, will not display the 'traditional' scholarship of the medical historian who has painstakingly examined each and every relevant ancient manuscript pertinent to his chosen field.

Indeed, my objective is to show that so-called traditional scholarship is an exercise that is 80% drudgery and 20% intellectuality. To write history, today as in the past, one must be capable of martyr-like perseverance. It is a back-breaking chore to identify and obtain suitable library materials. One of my library professors at Columbia University once said that the availability of a comprehensive citation index would probably abort 90% of the dissertations in the humanities and social sciences. My purpose is to show that he was correct to the extent that many dissertations are awarded as a sign of completing the monastic sentence of years of toil in the stacks of libraries.

When I agreed to speak, I wrote the Secretary-General that I would use the occasion to report to the medical library profession certain basic ideas I had first reported three years ago at the Symposium on the Foundations of Access to Knowledge (Garfield, 1968) in a paper entitled '"World Brain" or "Memex?" Mechanical and intellectual requirements for universal bibliographic control'. In spite of the essential novelty of these ideas for most of you, I could not, however, in clear conscience merely paraphrase or parrot material that is three years old. This would be disrespectful to the importance of such an international conclave. I will, therefore, limit my initial remarks to a brief presentation of the basic notions involved in comparing primordial citations, subject indexing, and historio-bibliography. I will then present some interesting new data generated since my first public discussion of primordial citations. Not the least of this is a list of the 50 most frequently cited journal articles and a recently compiled history of DNA updated since I first reported the history of the genetic code using citation analysis (Garfield *et al.*, 1964).

The appearance of the first 'experimental' *Science Citation Index* in 1963 created a mild furor in the literature. Not all the reviews were unfriendly: Professor Steinbach (1964), using a group of graduate students to help him review the *SCI*, said in *Science*:

> *Any real evaluation of* Science Citation Index *must be based on an extensive use test, and there has not been time for that. Most of us are accustomed to literature searches*

that begin with a subject. This, of course, presents real problems if one wishes complete coverage of the subject, because subject matter indexes are no better than the choice of words indexed. However, we are used to them — like an old shoe, they are comfortable.

On the other hand, a number of so-called reviews were in fact emotional and fearful responses to something quite different on the bibliographical scene — like a pair of new shoes. Most scientists and librarians, although working together on the frontiers of knowledge, are basically conservative. They are, after all, only human — and so am I. I can justify my own immodesty by referring to Professor Merton's recent AAAS paper (1969) in which he states that a scientist need not hide his vanity because it is quite healthy. The negative acclaim the *SCI* received by experts such as Cleverdon (1964) only convinced me that the *SCI* would be recognized as a milestone in medical and scientific bibliography. Like the savants of the last century who proved that airplanes could not fly, citation indexes *should* not work. But they do! This is not to say that there is not plenty of room for improvement. I find it hard to predict what the supersonic version will be. Possibly the major contribution of the *SCI* is that it contains a truly up-to-date calendar year author index — the *Source Index*. The *Source Index* is valuable not only in the process of citation verification and search by author, but will eventually become the means for correcting thousands of author-introduced citation errors that plague librarians every day.

A major semantic difficulty in discussing library systems is caused by the practice among librarians and others, particularly physicians and engineers, of lumping together two distinct problems of information retrieval — *information recovery* and *information discovery* (Garfield, 1966). Most scientists use author catalogs to find books they know exist. This I call information recovery. In this sense, the English word 'retrieval' is similar to the French word *retrouver* 'to find again'. Scientists rarely use subject catalogs to *recover* books. Many librarians have, therefore, justifiably asked why we spend so much money creating them (Gore, 1966). On the other hand, it is known that scientists do make use of periodical indexes. Subject indexes facilitate the process of information discovery — finding what is not known at the outset to exist. When the *Science Citation Index* entered the bibliographic scene, it added another means for accelerating information discovery. It is no surprise that the *SCI* appealed, at first, primarily to the adventuresome scholar who uses all sorts of serendipitous devices (Lederberg, 1959; Smith, 1964; Stonehill, 1965). This type of man is usually glad to discover the unexpected.

At first the librarian found *SCI* somewhat alien. Not only does a page of the *Citation Index* look strange (it could not have been otherwise), but the results of a search often seem equally strange. One cannot evaluate the results of many *SCI* discovery searches in exactly the same way that one can evaluate the traditional tool for information recovery. In retrospect, therefore, it is equally understandable that one of the major uses by librarians of the *Citation Index*, for which it was not designed, is citation verification. The intuition of the medical librarian on this is justified. In the seven years for which we now have citation indexes, an incredibly large percentage of the *entire* medical literature has been cited. There is a high but varying probability that, depending upon the year in which the paper was published, the citation one is at-

tempting to verify will be found in the *SCI*. Of 2,000,000 items cited in 1968 *alone*, about 25% or half a million were published in 1966 and 1967. This would account for a very substantial percentage of the items indexed in *Index Medicus*, *Chemical Abstracts*, *Biological Abstracts*, and *Excerpta Medica* combined. More importantly, it is as a tool for information discovery that the *Citation Index* section of the *SCI* must be evaluated. Regrettably, we do not have any established criteria for such measurement. Just as beauty is said to be in the eyes of the beholder, relevance is a quite subjective variable for the bibliographic explorer. What is relevant to one investigator is irrelevant to another.

One can develop methods for studying the overall retrieval effectiveness of the *SCI* and other indexes in well-defined search topics. For an extensive literature search on Thalidomide, Spencer (1967) compared the time of search with *SCI* to *Chemical Abstracts* and *Index Medicus*. Though favorable to *SCI*, such studies, however, have not revealed *why* the *SCI*, depending upon the circumstances, may or may not be very effective at all. Of course, we can conduct user evaluations in which users express general satisfaction or dissatisfaction, but this does not necessarily help us understand the fundamental conceptual problem of subject analysis.

To understand what is being retrieved in an *SCI* search, we have to recognize the underlying concept which is merely symbolized by a bibliographic citation. As librarians, our traditional concept of a 'subject' is so ingrained that we fail to realize that a word is merely a symbol for a concept. Chemists fall into the same trap and often forget that a chemical formula is only symbolic of the 'real' thing. Words, formulas, and citations are approximations. Furthermore, semanticists know that no two occurrences of the same word or symbol are identical. A subject heading or a key word functions as an approximation which is usually about one order of magnitude less specific than the approximation made by using a bibliographic citation as an indexing term. Citation indexing is not only 'in-depth' indexing as contrasted to the 'in-breadth' indexing of permuterm indexes, but the type of unique specificity the citation index provides is, at times, alarming to the traditional searcher. Indeed, a completely negative result in searching the indexes for current references to a particular paper or book may be exactly what the user expects or wants. Unfortunately, we have no standard of comparison for evaluating indexing systems in this respect.

To evaluate the specificity of citation indexing, one must translate a citation search question from the language of the citation index into the language of the word index. This is not easy, but when the attempt is made one recognizes that, as an indexing language, citation indexing also exhibits the characteristics of other indexing languages. For example, the *see* references and *see also* references contained in a typical controlled thesaurus can also be incorporated into citation indexes. As we will see later, in order to bridge the gap between the two indexing languages, I developed the concept of the primordial term — including primordial citations and primordial words.

One might ask why the term 'key citation', by analogy to 'key word', was not chosen. When I first used the noun phrase 'primordial citation' (Garfield, 1968), it was my intention that we design a dictionary of key citations. The dictionary would enable the librarian or student to make the transition from the symbolism of words

RANK	TOTAL TIMES CITED	AUTHOR	JOURNAL	VOL	PAGE	YEAR
1	2383	LOWRY OH	J BIOL CHEM	193	265	51
2	1964	REYNOLDS ES	J CELL BIOL	17	208	63
3	1461	LUFT JH	J BIOPHYS BIOCHEM CY	9	409	61
4	1318	FISKE CH	J BIOL CHEM	66	375	25
5	1167	FOLCH J	J BIOL CHEM	226	497	57
6	1066	BRAY GA	ANAL BIOCHEM	1	279	60
7	1000	SABATINI DD	J CELL BIOL	17	19	63
8	981	SPACKMAN DH	ANAL CHEM	30	1190	58
9	964	GORNALL AG	J BIOL CHEM	177	751	49
10	933	LINEWEAVER H	J AMER CHEM SOC	56	658	34
11	906	BURTON K	BIOCHEM J	62	315	56
12	775	DUNCAN DB	BIOMETRICS	11	1	55
13	774	SCHEIDEGGER JJ	INT ARCH ALLERGY APP	7	103	55
14	741	DOLE VP	J CLIN INVEST	35	150	56
15	726	DAVIS BJ	ANN NY ACAD SCI	121	404	64
16	723	NELSON N	J BIOL CHEM	153	375	44
17	723	REED LJ	AMER J HYG	27	493	38
18	718	MOORHEAD PS	EXP CELL RES	20	613	60
19	717	MARMUR J	J MOL BIOL	3	208	61
20	707	JACOB F	J MOL BIOL	3	318	61
21	703	WATSON ML	J BIOPHYS BIOCHEM CY	4	475	58
22	707	PALADE GE	J EXP MED	95	285	52
23	682	KARNOVSKY MJ	J BIOPHYS BIOCHEM CY	11	729	61
24	682	MARTIN RG	J BIOL CHEM	238	1372	63
25	675	SMITHIES O	BIOCHEM J	61	629	55
26	663	BARTLETT GR	J BIOL CHEM	234	466	59
27	652	BARKER SB	J BIOL CHEM	138	535	41
28	650	EAGLE H	SCIENCE	130	432	59
29	656	ROSENFELD AH	REV MOD PHYS	39	1	67
30	656	GELLMANN M	PHYS REV	125	1067	62
31	653	TREVELYAN WE	NATURE LOND	166	444	50
32	640	WARREN L	J BIOL CHEM	234	1971	59
33	640	ANDREWS P	BIOCHEM J	91	222	64
34	639	MONOD J	J MOL BIOL	12	88	65
35	636	SCHMIDT G	J BIOL CHEM	161	83	45
36	634	BARDEEN J	PHYS REV	108	1175	57
37	634	DEDUVE C	BIOCHEM J	60	604	55
38	631	KARPLUS M	J CHEM PHYS	30	11	59
39	631	AHLQUIST RP	AM J PHYSIOL	153	586	48
40	630	DUBOIS M	ANAL CHEM	28	350	56
41	628	ELLMAN GL	ARCH BIOCHEM BIOPHYS	82	70	59
42	625	WARBURG O	BIOCHEM Z	310	384	41
43	625	GELLMANN M	PHYSICS	1	63	64
44	624	MANDELL JD	ANAL BIOCHEM	1	66	60
45	623	DOLE VP	J BIOL CHEM	226	2905	57
46	622	LITCHFIELD JT	J PHARMAC EXP THER	96	99	49
47	622	MILLONIG G	J APPL PHYSICS	32	1637	61
48	618	FRIEDEMANN TE	J BIOL CHEM	147	415	43
49	618	MOORE S	J BIOL CHEM	211	907	54
50	618	JAFFE HH	CHEM REV	53	191	53

Fig. 1. Fifty most cited articles for 1967, ranked according to total times cited. (Refer to Appendix A)

to the symbolism of citations. Ordinarily, the subject expert does not require this assistance. The dictionary of key citations, however, soon became the *dictionary of primordial citations* for several reasons which are discussed below. But first I wish to note that a major portion of the work on this dictionary has now been completed as we have thus far compiled lists of the 20,000 most frequently cited papers for a five-year period. In Fig. 1, I have provided the list of 50 papers most frequently cited in the scientific literature during 1967. (See Appendix A for the titles of these papers.) Although I will not comment in detail on each paper, I do want to point out that many of these particular papers are methodological. In retrospect, one *expects* that such method papers will be frequently cited, but it comes as a surprise that they predominate so strongly. Furthermore, the *age* of these papers is even more dramatic, illustrating how today's research still depends upon methods and theories developed in previous generations. While examining the list of 'super-classics', as Professor Price (1965) would call them, one notices that the theoretical and other fundamental discovery papers also appear on the list. As we will see later, papers like these can be identified with the key events in the history of science or medicine. The predominance of biologically-oriented papers in contrast to those in the physical sciences is, of course, not a measure of the relative 'importance', social or otherwise, of molecular biology as contrasted to solid state physics. It probably simply reflects the quantitative differences in and character of publication in these areas.

But why is it not possible to construct a dictionary of *key* citations? Why a dictionary of *primordial* citations? We can, of course, in many cases associate a key

word with a key paper. The neologism 'euphenics', first used by Lederberg in 1963, can, of course, be used as a cross-reference to that paper. The underlying *concept* of euphenics, however, was known long before that time.

Many primordial citations identify key medical discoveries although, at the time of the discovery, an appropriate nomenclature was not even available. Consider the classical case of diabetes and the discovery of insulin by Banting and Best (Fig. 2).

A. BANTING, F. G. and BEST, C. H. (1922), Pancreatic extracts. *J. Lab. clin. Med.*, 7, 464.
B. BANTING, F. G. (1925), Nobel Prize Lecture.
C. BANTING, F. G., BEST, C. H. and MACLEOD, J. J. R. (1922), The internal secretion of the pancreas. *Amer. J. Physiol.*, 59, 479.
D. BANTING, F. G. and BEST, C. H. (1922), The internal secretion of the pancreas. *J. Lab. clin. Med.*, 7, 251.
E. BANTING, F. G., BEST, C. H., COLLIP, J. B., MACLEOD, J. J. R. and NOBLE, E. C. (1922), The effect of pancreatic extract (insulin) on normal rabbits. *Amer. J. Physiol.*, 62, 162.
F. BANTING, F. G., BEST, C. H., COLLIP, J. B., MACLEOD, J. J. R. and NOBLE, E. C. (1922), The effects of insulin on experimental hyperglycemia in rabbits. *Amer. J. Physiol.*, 62, 559.
G. BANTING, F. G., BEST, C. H., COLLIP, J. B., CAMPBELL, W. R. and FLETCHER, A. A. (1922), Pancreatic extracts in the treatment of diabetes mellitus, *Canad. med. Ass. J.*, 12, 141.
H. SCHMIDT, J. E. (1959), *Medical Discoveries (Who and When)*, p. 237. Thomas, Springfield, Ill.
I. SKINNER, H. A. (1961), *The Origin of Medical Terms*, p. 228. Williams and Wilkins, Baltimore.
J. DEMEYER, J. (1908), Glycolyse, hyperglycemic, glycosuric et diabete. *J. Méd. Brux.*, 13, 778.
K. BEST, C. H. (1960), Epochs in the history of diabetes. In: R. H. Williams (Ed), *Diabetes*, p. 1. Harper and Row, New York.
L. BEST, C. H. (1963), In: C. H. Best (Ed) *Selected Papers of Charles H. Best*. Univ. of Toronto Press, Toronto.

Fig. 2. Bibliography on insulin (Banting and Best).

The association between diabetes mellitus and pancreatic defect was known for nearly 30 years prior to the discovery of insulin. In a historical review (A), Banting and Best refer to an early success by George Ludwig Zuelzer, a German physician who isolated a crude pancreatic extract in 1908. This Zuelzer used to treat diabetes in several patients and some improvement was noted. Unpredictable side reactions and failure by others led to abandonment of this treatment. Until then diabetic control had been limited to carbohydrate deprivation. The dietetic approach eventually produced starvation, overwhelming infection, coma, and death. As shown in Fig. 2, the first hint of their historic discovery, according to Banting's Nobel Prize lecture (B), appears in the December 1921 *Proceedings of the American Physiological Society*. This report was later abstracted and expanded in two journal articles in 1922 (C, D) under the title 'Internal secretion of the pancreas'. The word 'insulin' was not used. In another research paper (E) which followed, however, the word 'insulin' does appear in the title but in parenthesis after the expression 'pancreatic extract'. In a research paper subsequently published (F), the word 'insulin' is used and 'pancreatic extract' is omitted.

Banting and Best do not give their reason for coining the word. The point I wish to stress is that the first case report of the clinical use of insulin which is often cited as a classic (G) did *not* contain the word 'insulin'.*

'Insulin' first appears as a main index word in the 1923 2nd Quarter *Index Medicus*.

* In extensively reviewing medical histories, Best's memoirs, etc. (K, L), my colleague, Dr.

Gene Joslin mentions a 1921 notebook of Best in which the word 'isletin' is used and that Banting and Best used the word 'insulin' orally two months after publication of the classical 1922 paper published in the *Canadian Medical Association Journal.*

The important point I am trying to stress in this typical example of what structural linguists call the process of *analogous linguistic change* is that primordial citations must be distinguished from primordial words. Only an *a posteriori* intellectual effort can clearly identify what might then be called a 'key' citation. For any student who wants a quick identification of the classical paper on the clinical use of insulin mentioned above, *The Dictionary of Primordial Citations* will be extremely useful. The reverse may also be true. The paper or book with which a concept may become identified may appear many years after the term is in vogue or being heavily used. In fact, many times no clearly identifiable citation is associated with the word. As any etymologist knows, to identify the first occurrence of a word or phrase is no small task; and each particular subsequent use, whether in lay usage or in scientific usage, is only a shade different than the previous use.

To amplify the difficulties in correlating complex concepts with traditionally word-structured indexing languages, consider the concept 'protein determination by the Folin phenol reagent', sometimes referred to as the 'Lowry method'. In Fig. 1, we saw that this was first reported in 1951 and the paper is the most frequently cited work in the 1967 literature. No term for it exists in the *Medical Subject Headings List* (MeSH) of *Index Medicus*. The symbol **Lowry 1951 JBC,** however, adequately identifies the concept. The symbol **Lowry 1951, JBC vol. 193, p. 265** also identifies its exact address! Unquestionably, *Index Medicus* does provide for indexing papers on protein determination methods, but that is a vastly more generic concept than the Lowry method or derivates thereof.

Perhaps this does not seem particularly important in a medical index, but does it seem unreasonable that a researcher might ask for papers in which the Lowry method has been employed in cancer research? From the number of papers on this topic alone, one must conclude that the depth of indexing this implies is necessary, and further, we must find ways to bridge the gap between citation indexes and word indexes. *The Dictionary of Primordial Citations* can help resolve some of these problems, but must be limited to those citations which by definition have become classics. We can only hope to develop the word synonyms or equivalents for each of about 20,000 of the most frequently cited papers each year – about 1% of all the papers that are cited. Should we attempt to establish key or primordial citations for those older words or word phrases which occur most frequently? Clearly, this is an entirely

---

Richard Torpie of Hahnemann Medical College was unable to find mention of the decision to use the word 'insulin'. Schmidt (H), however, ascribes to Jean de Meyer, a French physiologist, the term *insuline, circa* 1909. Skinner (I) reminds us that the word 'insulin' is a derivative of the Latin *insula* 'island'. Of course, the active ingredient is derived from the Islands or Islets of Langerhans of the pancreas. De Meyer states that it was Schaefer who presupposed in 1913 that the Islands of Langerhans were responsible for the active principle long before the extract was obtained. Banting, Best, and MacLeod isolated the substance in Toronto in 1921 and used the name 'insulin' for their extract. We could not locate any article by Schaefer; de Meyer, however, did write on the subject of diabetes (J). Of significance, too, is the methodical citation by Banting and Best of Langerhans' discovery in all their early work.

different and possibly futile exercise. Frequency of word usage in scientific titles or traditional indexing languages is not going to provide a necessarily useful approach to the current literature. The historian would have great interest in knowing the primordial citations for words like 'cancer', 'liver', etc., but the searcher interested in some specialized aspect of cancer or liver research would not be aided significantly by such devices. In any case, extremely useful by-products can be obtained from large-scale word-frequency analyses. Before discussing these, let me cite a current example which illustrates why citation language is essential to current information retrieval.

Suppose that a physician comes to your library and requests current information on the 'Chinese Restaurant Syndrome'. This might seem like a jest, but in fact just last year it was discussed in the *New England Journal of Medicine* (Schaumburg *et al.*, 1968) and later in *Science* (Schaumburg *et al.*, 1969). The topic has also been discussed recently in the *New Scientist* under the dubious heading of 'Kwok's disease' (Chedd, 1969). These reference citations will continue to be useful citation index headings to help scientists retrieve information on this topic. But how will the medical librarian bridge the gap between the terms 'CRS' or 'Kwok's disease' and these primordial citations? We were acutely conscious of this gap between the indexing language of the citation index and the natural language of science when we introduced the concept of permuterm indexing.

The *Permuterm Subject Index* section of the *SCI*, which is still relatively unknown to many medical librarians, is based upon title words. *PSI* is obviously related to the Key-Word-in-Context (KWIC) index which has become so widely known through its use in *Biological Abstracts* and *Chemical Titles* (Luhn, 1959). Since KWIC and KWOC — or Key-Word-Out-of-Context index, not to be confused with Kwok's disease — are both title-derived, there are certain similarities between them and *PSI*. Their differences, however, are equally significant.

CONVENTIONAL PERMUTERM	MODIFIED PERMUTERM
BIRTH CONTROL GROWTH POPULATION RATE	BIRTH-CONTROL POPULATION-GROWTH BIRTH-RATE
CONTROL BIRTH GROWTH POPULATION RATE	
	POPULATION-GROWTH BIRTH-CONTROL BIRTH-RATE
GROWTH BIRTH CONTROL POPULATION RATE	
POPULATION BIRTH CONTROL GROWTH RATE	
	BIRTH-RATE POPULATION-CONTROL BIRTH-CONTROL
RATE BIRTH CONTROL GROWTH POPULATION	

Fig. 3. Permuterm indexing of 'Control of population growth and birthrate'.

In the *Permuterm Index* every significant title word is *permuted*, not merely rotated as in KWIC, to produce all possible pairs of terms. Thus, approximately $n(n\text{-}1)$ term pairs are created by this procedure. In a title containing six significant words, thirty pairs are created; for five terms, twenty pairs are created.

In very recent work we have developed modified permuterm computer programs which automatically or algorithmically generate 'logical' subdivisions in an index. This approach, like our studies of citation frequency, is based on purely quantitative measures of word co-occurrences. These frequency analyses establish *semantically* useful word phrases and word pairs. Such analyses should not be confused with textual word-frequency studies. We have recently completed a statistical analysis of several million word and word-phrase occurrences for the 300,000 titles appearing in the 1967 *SCI Source Index*. These titles are the initial input for the *Permuterm Subject Index*.

It is important to observe that when one seeks information on a highly specific topic, it makes very little difference, except for *format* considerations, whether or not he uses a KWIC or a permuterm index. If only one or two articles are identified in any system, then one can quickly scan the article title. Most scientists *reject* KWIC indexes precisely on the grounds of format. Secondly, and more importantly, when one searches a subject for which there are dozens of articles, one needs subdivisions to narrow the search to a few pertinent items. This is largely achieved in the format of the *PSI*. But the pure permutation of significant title words does not contend with the peculiar word or noun-phrase constructions of the English language. This is sometimes aggravated by omission of punctuation marks. Thus, consider the importance of the comma in the sentence, 'Doctor X, while distilling alcohol, was consumed'. Contrast this to 'Doctor X, while distilling, consumed alcohol' and 'Doctor X consumed distilling alcohol'. 'Distilling' and 'distilling alcohol' are quite distinct semantic concepts and ideally one wishes to preserve such distinctions. In an index one may sacrifice such distinctions to increase overall retrieval effectiveness and indexing economy.

How exciting to find that, by large-scale statistical analysis, the frequency of such unwanted co-occurrences is limited to an extremely small number. If one establishes a minimum threshold of co-occurrence, then legitimate word phrases are identified. If two consecutive words occur in titles x or more times, then that word pair has been established as a legitimate word phrase. Thus, while 'distilling alcohol' might in fact occur only once or twice, *if* the sequence *did* occur ten times, it would prove to be a useful primary indexing term! This seemingly innocuous discovery has great significance for the efficient design of indexes, since we can now reduce the number of permutations while *increasing* retrieval speed and specificity.

Consider the indexing of 'Control of population growth and birthrate' (Fig. 3). Whereas a concept like 'birth control' would appear as two primary terms by pure and simple permutation, the procedure described above *automatically* indexes this title under **birth-control**. Unfortunately, the procedure is not all that simple because we do not wish to separate the term 'birth-control' from 'control-of-birth'. It is precisely with this in mind that one must perform the frequency analyses after the permutation process and then reassign the indexing terms once the appropriate word

pairs have been identified. This procedure resolves the problem of conjunctive phrases in which one finds expressions such as 'control of population growth and birth rate'. By the procedure I have just described, such an article will be indexed under **birth-control, birth-rate, population-growth,** etc., whereas previously, the primary terms would be **birth, control, growth, rate.** In other words, the computer first examines the twenty word pairs created by permutation and replaces the single-term entries by the hyphenated expressions once it is determined that the word pair occurs above a given threshold.

Fig. 3 shows the indexing terms which would result from the second procedure, depending upon the statistics one might find for a particular file of information. All high-frequency term pairs would be cross-referenced to the appropriate term since they now function as primary terms. Thus, **control-growth** would be cross-referenced to **birth-control.** All such studies, of course, accentuate the advantages that may be derived from pre-edit and post-edit procedures by human editors who can perform the important indexing function of suppressing useless indexing entries. Using procedures of this kind, in the future, monitoring the changing literature of science and medicine will be possible by whatever quantitative criteria one wishes to select. One can establish useful word phrases without resorting to human editing procedures.

It is essential to keep in mind that the deliberate purpose of the *Permuterm Index*, and indeed most co-ordinate indexing systems, is to direct the reader quickly to a small set of references. Whenever the reader finds more than ten articles indexed under a given primary term, we must provide him further means of refining his search. It should also be remembered that the *PSI* was expressly designed to augment the *Citation Index*, to foster information *recovery* for a partially remembered title when a key word is known but not a citation.

In a similar fashion, we have established that the occurrence of a given reference citation 15 or more times in a given year clearly identifies a *putative* primordial term which should be characterized in natural-language terms for our *Dictionary of Primordial Citations*. We must realize that this is a constantly changing task. The Banting and Best paper on pancreatic extract mentioned earlier would be sought under the term **insulin.** The searcher wants mechanisms for quickly identifying reasonable numbers of references in a reasonable time. Dictionaries or thesauri based on these frequency analyses appear to be reasonable objectives. Of course, this can also be done with a controlled authority list like MeSH. But changes in MeSH result from analysis of indexing practices rather than analysis of the terminology occurring in the medical literature. There is no reason, however, why the two approaches cannot eventually be reconciled.

I would now like to turn from the theory of bibliographic symbols to the field of historio-bibliography. If I may paraphrase a great American, Dr. Martin Luther King, I have a dream. In Wellsian terms, this dream was symbolized as *World Brain* and by Vannevar Bush (1945) as *Memex*. Unlike Mr. Wells, I hope to see *my* dream become a reality while I am still among you.

In the first part of my presentation, I discussed the primordial term as it related to the traditional problem of subject analysis of library materials. At least one major significant by-product is attached to the use of primordial citations, which in this

(continued)

respect differ from their counterpart, primordial words. Bibliographic citations, as we have seen, not only identify or symbolize subject matter, but as 'addresses', citations contain chronological information which permit one to easily arrange them. When this is done, one has a crude history of the development of a subject. This is not new. Retrospective bibliographies have been arranged in chronological order for quite some time. But now, let us see what happens when we use, not merely the citations which identify the source documents, but also the reference citations. In Fig. 4, I have drawn a circle for each citation shown in a bibliography on staining of nucleic acids, and given each one an accession number. Unlike a traditional bibliography, the set of 15 source citations is drawn in a network diagram in which the lines with arrows

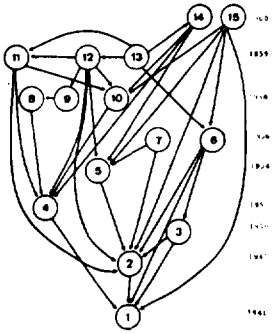

Fig. 4. Citation network of articles on nucleic acids. Citation relationships illustrated by network of 15 papers from a bibliography on nucleic acid staining.

indicate that, for example, paper 13 has cited paper 6. Anyone can create such a diagram for a simple network and I always make my students at the University of Pennsylvania do this when they compile a bibliography. When the number of source documents in the network becomes quite large, however, one can run into considerable difficulty in simply portraying this information. In a recent paper we have shown how these problems of display can be overcome (Garfield and Sher, 1967; *see also* Garfield and Malin, 1969). It is not my intention or purpose to digress to this interesting problem. The important point I wish to stress is that we have available a means for displaying citation networks without human intervention.

What is the significance of all this for the medical historian and bibliographer? It means that, in the near future, the compilation of bibliographies will be inseparable from writing the history of that field. A scholar will be able to sit before his computer console and he will specify some starting point — a person, a word, a citation, a place. Given a particular word or document, he will then ask the computer to display a list of pertinent papers. Then the computer will draw or display for him a historical road map which will show him not merely the list of papers and books, but also a graphical approximation or detailed history of that subject. In an earlier paper (Garfield *et al.*, 1964), we simulated this process by reconstructing the recent history of the genetic code by a process of citation analysis. At that time we traced the history up to the time of Nirenberg's now classical paper.

It is difficult to comprehend how hard it is to display such information until one tries to draw the *complete* diagram of any given field. But again, frequency analysis simplifies the problem; with certain exceptions we can eliminate anything from the overall network which does not satisfy a given critical threshold of citation linkage, and place it *temporarily* in a computer storage area. When we wish to examine the particular period in history more closely, we can do so by zooming in, and then, as historians, try to understand what significance, if any, some of the many uncited papers may have. We know, in fact, that probably 10% or more of the literature is never cited again once it is published – possibly a measure of the redundancy necessary to insure that any average paper does, in fact, get into the general stream of things (Price, 1965).

The recent history of DNA was reconstructed by vastly more simple procedures than that which we employed to do the early history of the genetic code. The basic assumption was simple: given a list of the recent papers on the topic, about 30 or 40 published in 1967 and cited in a single review or found in a straightforward literature search, the bibliographies of all the 1967 papers were examined and a master list compiled. Since several hundred papers were cited, all were eliminated which were cited only once. By a process of iteration, the next group of cited references to be eliminated were those cited only twice, etc. Eventually, this led to the list of papers shown in Fig. 5, each of which was cited five or more times. Subsequently, the list of papers was checked in the 1967 *Science Citation Index* and we attained a further verification of the significance of each paper by ascertaining that they are also highly cited in general. It is significant that for a fast-moving, active field like molecular biology, one must repeat this type of procedure for each preceding year if one wishes to completely fill in the eventful years from 1961 to 1967, during which time we have come from the breaking of the genetic code all the way to *in vitro* synthesis of life in the recent work of Kornberg *et al.* (refer to Appendix B for citation data to Fig. 5).

Of further significance is that *many* of these papers (indicated by black circles in Fig. 5) appear on our list of most heavily cited papers in the literature. Since that list is confined to the 1% per year which are cited 15 or more times per year, one would expect that a lower rate, about 5 cites per year as it turns out, would be sufficient for a specialized field. Thus, to write the entire history of science and medicine as distinct from merely writing the history of DNA or any other specific topic, one's interest would center on events of broader impact and scope.

By way of reiteration, I wish to mention that this history of DNA was written by my assistant, Marie V. V. Williams, under my instructions, even though neither of us knows anything about genetics. I do not think any geneticist would seriously challenge the diagram in Fig. 5, and it, therefore, becomes a perfectly valid teaching aid to the student and a great time saver for the historian.

Let me spell out the implications of these examples – if they are not self-evident from my discussion – for your future dealings with the reader who is faced with a common problem: given a bibliography of 100 papers on any selected field – and today that is commonplace – how can one select the key group of papers to read *first?* One must make choices since he cannot possibly read everything. Here you have seen how, starting with several hundred references, we have identified a dozen or so

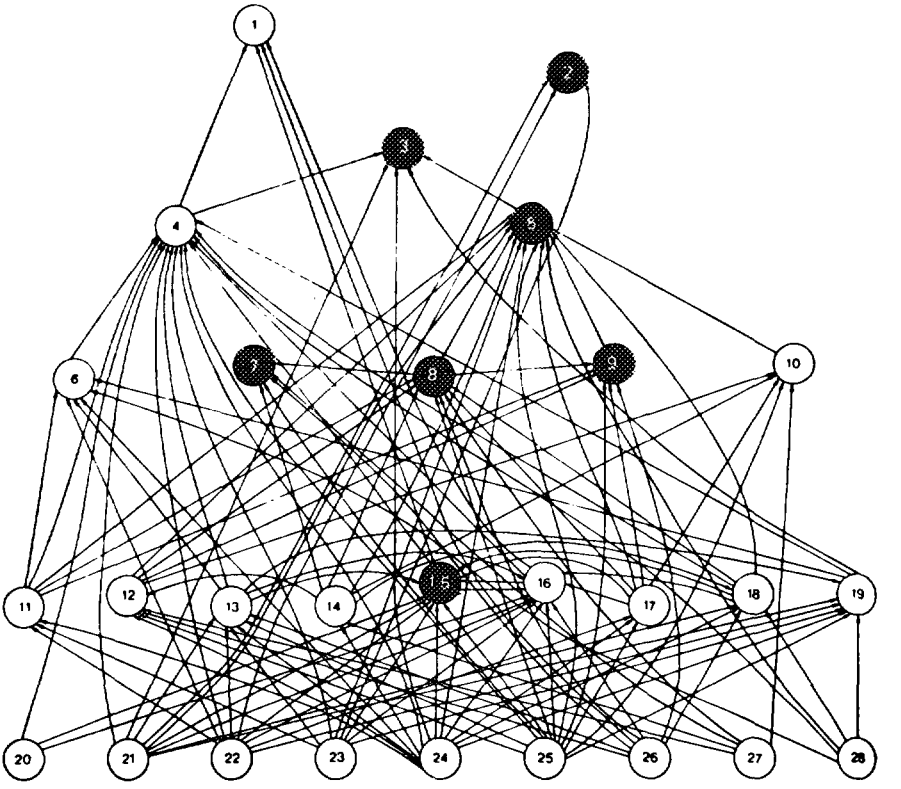

Fig. 5. Citation network of DNA articles based on review of 1967 literature by A. Sadgopal in *Advances* in *Genetics* (Academic Press, New York, 1968, v. 14, p. 325-404). Legend (refer to Appendix B): 1, Sheehan 1958; 2, Bray 1960; 3, Nirenberg 1961; 4, Marcker 1964; 5, Nirenberg 1964; 6, Marcker 1965; 7, Brenner 1965; 8, Khorana 1965; 9, Nirenberg 1965; 10, Khorana 1965; 11, Marcker 1966; 12, Khorana 1966; 13, Marcker 1966; 14, Khorana 1966; 15, Adams 1966; 16, Webster 1966; 17, Nirenberg 1966; 18, Ochoa 1966; 19, Nakamoto 1966; 20, Berberich 1967; 21, Lucas-Leonard 1967; 22, Caskey 1967; 23, Ochoa 1967; 24, Khorana 1967; 25, Nirenberg 1967; 26, Ochoa 1967; 27, Khorana 1967; 28, Ochoa 1967.

RANK	AUTHOR	TOTAL TIMES CITED	RANK	AUTHOR	TOTAL TIMES CITED
1	LOWRY OH	2021	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	711
5	PAULING L	1053	30	BORN M	710
6	GELLMANN M	947	31	BRACHET J	708
7	COTTON FA	940	32	RINGSTEIN E	702
8	POPLE JA	933	33	ALBERT A	687
9	BELLAMY LJ	908	34	LUFT JH	674
10	SNEDECOR GW	904	35	DEDUVE C	673
11	BOYER PD	883	36	VONEULER US	668
12	BAKER BR	878	37	FIESER LF	666
13	KOLTHOFF IM	863	38	HUISGEN R	661
14	HERZBERG G	842	39	MOVIKOFF AB	656
15	FISCHER F	828	40	GOODWIN TW	643
16	SEITZ F	822	41	BARTON DHR	632
17	DJERASSI C	801	42	FISHER RA	631
18	BERGMEYER HU	794	43	BATES DR	627
19	WEBER G	780	44	FLORY PJ	626
20	REYNOLDS ES	748	45	STAHL E	626
21	MOTT NF	741	46	DEWAR MJS	618
22	ECCLES JC	737	47	GILMAN H	618
23	FRIGL F	736	48	FOLCH J	616
24	FREUD S	727	49	DISCHE Z	614
25	PEARSE AGE	726	50	GLICK D	608

Fig. 6. Fifty most cited authors for 1967, ranked according to total times cited.

papers which represent the core of this field, and the 'field' can, of course, be individually tailored to the reader's needs. If you have done the recent history of DNA for one student, it can be used by another; but if faculty members or researchers have chosen less known topics, one must be equally prepared to solve their selective reading problems as well.

Finally, let me briefly turn from the topic of historio-bibliography to that of sociology. At the recent AAAS meeting I presented a paper, 'Can Nobel Prize winners be predicted?' (Garfield and Malin, 1968). The title was somewhat facetious, but actually a more correct title would be 'Can the Nobel Prize winners be forecasted?'. 'To predict' is a very strong term, one expected from the followers of Nostradamus. 'To forecast' is a probabilistic term: a meteorologist forecasts the weather by stating certain probabilities; he cannot predict the weather with absolute certainty.

In the same way, it is not possible to predict using the *SCI*; it is possible, however, to say that from the list of men shown in Fig. 6, one can forecast with high probability that several will receive the Nobel Prize. This is no small achievement 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.

The ultimate decisions will, of course, be made by their peers in the Swedish Academy, etc., but there can be little doubt, as was stated by Newell (1962), that citation indexes will be used increasingly as a means of evaluating scientific merit. This was originally proposed by Golay (1953) and recently expressed by Cranberg (1969) in *Physics Today*. This will, of course, require more meticulous attention to bibliographic practices to insure fair treatment for all, but within the bounds of acceptable error, the evidence is very clear that the *SCI* has become a major sociometric tool. The recent work of the Coles (1968) and others is merely a harbinger of future developments.

I have tried to show the inseparable relationship that exists between the conceptual problems of bibliographic control, subject analysis, symbol theory, and the history and sociology of medicine. It has been an ambitious undertaking. Undoubted-

ly, I have only scratched the surface and I leave it to others with less pragmatic concerns than publishing a work of the size and scope of the *SCI*. Let the scholars like Professors Merton and Price do their job. We have certainly given them all the ammunition they need.

In closing, let me relate that we now plan to complete the data base that will be needed to fully arm the historian who wishes to deal with the history of the decade 1961-1970. As soon as practical, we will fill in the *SCI* for the missing years of 1962 and 1963, and at the same time use the ten-year data base to create discipline-oriented indexes which will include chemistry and physics as well as the social sciences and education. By the time this enormous data base is completed, we expect that our computer hardware and software will be caught up and the dream I have sketched here will be realized at least insofar as we presently conceive of it.

appendix A

Titles of fifty most cited articles for 1967 ranked according to total number of times cited (refer to Fig. 1).

rank

1. LOWRY, O. H., ROSEBROUGH, N. J., FARR, A. L. and RANDALL, R. J., Protein measurement with the folin phenol reagent.
2. REYNOLDS, E. S., The use of lead citrate at high pH as an electron-opaque stain in electron microscopy.
3. LUFT, J. H., Improvements in epoxy resin embedding methods.
4. FISKE, C. H. and SUBBAROW, Y., The colorimetric determination of phosphrous.
5. FOLCH, J., LEES, M. and SLOANE STANLEY, G. H., A simple method for the isolation and purification of total lipides from animal tissues.
6. BRAY, G. A., A simple efficient liquid scintillator for counting aqueous solutions in a liquid scintillation counter.
7. SABATINI, D. D., BENSCH, K. and BARRNETT, R. J., Cytochemistry and electron microscopy: the preservation of cellular ultrastructure and enzymatic activity by aldehyde fixation.
8. SPACKMAN, D. H., STEIN, W. H. and MOORE, S., Automatic recording apparatus for use in the chromatography of amino acids.
9. GORNALL, A. G., BARDAWILL, C. J. and DAVID, M. M., Determination of serum proteins by means of the biuret reaction.
10. LINEWEAVER, H. and BURK, D., The determination of enzyme dissociation constants.
11. BURTON, K., A study of the conditions and mechanism of the diphenylamine reaction for the colorimetric estimation of deoxyribonucleic acid.
12. DUNCAN, D. B., Multiple range and multiple F tests.
13. SCHEIDEGGER, J. J., A micro-method for immuno-electrophoresis. (In French).
14. DOLE, V. P., A relation between non-esterified fatty acids in plasma and the metabolism of glucose.
15. DAVIS, B. J., Disc electrophoresis. II. Method and application to human serum proteins.
16. NELSON, N., A photometric adaption of the Somogyi method for the determination of glucose.
17. REED, L. J. and MUENCH, H., A simple method of estimating fifty per cent endpoints.
18. MOORHEAD, P. S., NOWELL, P. C., MELLMAN, W. J., BATTIPS, D. D. and HUNGERFORD, D. A., Chromosome preparations of leukocytes cultured from human peripheral blood.
19. MARMUR, J., A procedure for the isolation of deoxyribonucleic acid from micro-organisms.
20. JACOB, F. and MONOD, J., Genetic regulatory mechanisms in the synthesis of proteins.
21. WATSON, M. L., Staining of tissue sections for electron microscopy with heavy metals.
22. PALADE, G. E., A study of fixation for electron microscopy.
23. KARNOVSKY, M. J., Simple methods for staining with lead at high pH in electron microscopy.

24. MARTIN, R. G. and AMES, B. N., A method for determining the sedimentation behavior of enzymes: application to protein mixtures.
25. SMITHIES, O., Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults.
26. BARTLETT, G. R., Phosphorus assay in column chromatography.
27. BARKER, S. B. and SUMMERSON, W. H., The colorimetric determination of lactic acid in biological material.
28. EAGLE, H., Amino acid metabolism in mammalian cell cultures.
29. ROSENFELD, A. H., BARBARO-GALTIERI, A., PODOLSKY, W. J., PRICE, L. R., SODING, P., WOHL, C. G., ROOS, M. and WILLIS, W. J., Data on particles and resonant states.
30. GELL-MANN, M., Symmetries of baryons and mesons.
31. TREVELYAN, W. E., PROCTER, D. P. and HARRISON, J. S., Detection of sugars on paper chromatograms.
32. WARREN, L., The thiobarbituric acid assay of sialic acids.
33. ANDREWS, P., Estimation of the molecular weights of protein in Sephadex gel-filtration.
34. MONOD, J., WYMAN, J. and CHANGEUX, J. P., On the nature of allosteric transitions: a plausible model.
35. SCHMIDT, G. and THANNHAUSER, S. J., A method for the determination of desoxyribonucleic acid, ribonucleic acid, and phosphoproteins in animal tissues.
36. BARDEEN, J., COOPER, L. N. and SCHRIEFFER, J. R., Theory of superconductivity.
37. DEDUVE, C., PRESSMAN, B. C., GIANETTO, R., WATTIAUX, R. and APPELMANS, F., Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue.
38. KARPLUS, M., Contact electron-spin coupling of nuclear magnetic movements.
39. AHLQUIST, R. P., A study of the adrenotropic receptors.
40. DUBOIS, M., GILLES, K. A., HAMILTON, J. K., REBERS, P. A. and SMITH, F., Colorimetric method for determination of sugars and related substances.
41. ELLMAN, G. L., Tissue sulfhydryl groups.
42. WARBURG, O. and CHRISTIAN, W., Isolation and crystallization of the fermentation ferment enolase. (In German).
43. GELL-MANN, M., The symmetry group of vector and axial vector currents.
44. MANDELL, J. D. and HERSHEY, A. D., A fractionating column for analysis of nucleic acids.
45. DOLE, V. P. and MEINERTZ, H., Microdetermination of long-chain fatty acids in plasma and tissues.
46. LITCHFIELD JR., J. T. and WILCOXON, F., A simplified method of evaluating dose-effect experiments.
47. MILLONIG, G., Advantages of a phosphate buffer for OsO_4 solutions in fixation.
48. FRIEDEMANN, T. E. and HAUGEN, G. E., Pyruvic acid. II. The determination of keto acids in blood and urine.
49. MOORE, S. and STEIN, W. H., A modified ninhydrin reagent for the photometric determination of amino acids and related compounds.
50. JAFFE, H. H., A reexamination of the Hammett equation.

appendix B

Citations to network of DNA articles based on review of 1967 literature by A. Sadgopal in *Advances in Genetics* (Academic Press, New York, 1968, v. 14, p. 325-404) (refer to Fig. 5).

node

1. SHEEHAN, J. C. and YANG, D. M. (1958), The use of N-formylamino acids in peptide synthesis. *J. Amer. Chem. Soc.*, 80, 1154.
2. BRAY, G. A. (1960), A simple efficient liquid scintillator for counting acqueous solutions in a liquid scintillation counter. *Analyt. Biochem.*, 1, 279.
3. NIRENBERG, M. and MATTHAEI, J. H. (1961), The dependence of cell-free protein synthesis in

E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. nat. Acad. Sci. (Wash.)*, 47, 1588.
4. Marcker, K. A. and Sanger, F. (1964). N-formylmethionyl-sRNA. *J. molec. Biol.*, 8, 835.
5. Nirenberg, M. and Leder, P. (1964). RNA codewords and protein synthesis-effect of trinucleotides upon binding of sRNA to ribosomes. *Science*, 145, 1399.
6. Marcker, K. (1965). Formation of N-formyl-methionyl-sRNA. *J. molec. Biol.*, 14, 63.
7. Brenner, S., Stretton, A. O. W. and Kaplan, S. (1965). Genetic code — nonsense triplets for chain termination and their suppression. *Nature*, 206, 994.
8. Soll, D., Ohtsuka, E., Jones, D. S., Lohrmann, R., Hayatsu, H., Nishimura, S. and Khorana, H. G. (1965). Studies on polynucleotides. 49. Stimulation of binding of aminoacyl-SRNAS to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc. nat. Acad. Sci. (Wash.)*, 54, 1378.
9. Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F. and O'Neal, C. (1965). RNA codewords and protein synthesis. 7. On general nature of RNA code. *Proc. nat. Acad. Sci. (Wash.)*, 53, 1161.
10. Nishimura, S., Jones, D. S., Ohtsuka, E., Hayatsu, H., Jacob, T. M. and Khorana, H. G. (1965). Studies on polynucleotides. 47. *In vitro* synthesis of homopeptides as directed by a ribopolynucleotide containing a repeating trinucleotide sequence — new codon sequences of lysine glutamic acid and arginine. *J. molec. Biol.*, 13, 283.
11. Bretscher, M. S. and Marcker, K. A. (1966). Polypetidyl-s-ribonucleic acid and aminoacyl-s-ribonucleic acid binding sites on ribosomes. *Nature*, 211, 380.
12. Jones, D. S., Nishimura, S. and Khorana, H. G. (1966). Studies on polynucleotides. 56. Further syntheses in vitro of copolypeptides containing 2 amino acids in alternating sequence dependent upon DNA-like polymers containing 2 nucleotides in alternating sequence. *J. molec. Biol.*, 16, 454.
13. Clark, B. F. C. and Marcker, K. A. (1966). N-formyl-methionyl-s-ribonucleic acid and chain initiation in protein biosynthesis — polypeptide synthesis directed by a bacteriophage ribonucleic acid in a cell-free system. *Nature*, 211, 378.
14. Morgan, A. R., Wells, R. D. and Khorana, H. G. (1966). Studies on polynucleotides. 59. Further codon assignments from amino acid incorporations directed by ribopolynucleotides containing repeating trinucleotide sequences. *Proc. nat. Acad. Sci. (Wash.)*, 56, 1899.
15. Adams, J. M. and Capecchi, M. R. (1966). N-formylmethionyl-sRNA as initiator of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 55, 147.
16. Webster, R. E., Engelhardt, D. L. and Zinder, N. (1966). In vitro protein synthesis — chain initiation. *Proc. nat. Acad. Sci. (Wash.)*, 55, 155.
17. Kellogg, D. A., Doctor, B. P., Loebel, J. E. and Nirenberg, M. (1966). RNA codons and protein synthesis. 9. Synonym codon recognition by multiple species of valine-, alanine-, and methionine-sRNA. *Proc. nat. Acad. Sci. (Wash.)*, 55, 912.
18. Stanley, W. M., Salas, M., Wahba, A. J. and Ochoa, S. (1966). Translation of genetic message — factors in initiation of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 56, 290.
19. Nakamoto, T. and Kolakofsky, D. (1966). A possible mechanism for initiation of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 55, 606.
20. Berberich, M. A., Kovach, J. S. and Goldberger, R. F. (1967). Chain initiation in a polycistronic message — sequential versus simultaneous derepression of enzymes for histidine biosynthesis in Salmonella typhimurium. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1857.
21. Lucas-lenard, J. and Lipmann, F. (1967). Initiation of polyphenylalanine synthesis by N-acetylphenylalanyl/sRNA. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1050.
22. Caskey, C. T., Redfield, B. and Weissbach, H. (1967). Formylation of guinea pig liver methionyl-sRNA. *Arch. Biochem.*, 120, 119.
23. Salas, M., Hille, M. B., Last, J. A., Wahba, A. J. and Ochoa, S. (1967). Translation of genetic message. 2. Effect of initiation factors on binding of formyl-methionyl-tRNA to ribosomes. *Proc. nat. Acad. Sci. (Wash.)*, 57, 387.
24. Ghosh, H. P., Söll, D. and Khorana, H. G. (1967). Studies on polynucleotides. 67. Initiation of protein synthesis in vitro as studied by using ribopolynucleotides with repeating nucleotide sequences as messengers. *J. molec. Biol.*, 25, 275.

25. MARSHALL, R. E., CASKEY, C. T. and NIRENBERG, M. (1967), Fine structure of RNA codewords recognized by bacterial amphibian and mammalian transfer RNA. *Science*, 155, 820.
26. LAST, J. A., STANLEY, W. M., SALAS, M., HILLE, M. B., WAHBA, A. J. and OCHOA, S. (1967), Translation of genetic message. 4. UAA as a chain termination codon. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1062.
27. KÖSSEL, H., MORGAN, A. R. and KHORANA, H. G. (1967), Studies of polynucleotides. 73. Synthesis in vitro of polypeptides containing repeating tetrapeptide sequences dependent upon DNA-like polymers containing repeating tetranucleotide sequences — direction of reading of messenger RNA. *J. molec. Biol.*, 26, 449.
28. SALAS, M., MILLER, M. J., WAHBA, A. J. and OCHOA, S. (1967), Translation of genetic message. 5. Effect of Mg^{++} and formylation of methionine in protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1865.

references

BUSH, V. (1945), As we may think. *Atlantic Monthly*, 176, 101.
CHEDD, G. (1969), Kwok's disease. *New Scientist*, 41, 504.
CLEVERDON, C. W. (1964), Citation indexing: Science Citation Index. *Nature*, 203/4944, 446.
COLE, S. and COLE, J. R. (1968), Visibility and the structural bases of awareness of scientific research. *Amer. Sociol. Rev.*, 33/3, 397.
CRANBERG, L. (1969), Citations and evaluation. *Physics Today*, 22/4, 15.
GARFIELD, E. (1966), ISI eases scientists' information problems, provides convenient orderly access to literature. *Karger Gaz.*, 13, 2. Reprinted (1969): *Science*, 154, 762.
GARFIELD, E. (1968), 'World Brain' or 'Memex'? Mechanical and intellectual requirements for universal bibliographic control. In: E. B. Montgomery (Ed) *The Foundations of Access to Knowledge*, p. 169. Syracuse University Press, New York.
GARFIELD, E. and MALIN, M. V. (1968), Can Nobel Prize winners be predicted? Presented at AAAS Meeting, Dallas, Texas, December 28, 1968.
GARFIELD, E. and MALIN, M. V. (1969), Diagonal Display — a new technique for graphic presentation of network diagrams. In: 1969 *Transactions of the American Association of Cost Engineers, Thirteenth National Meeting, Pittsburgh, Pennsylvania*, p. 222, (Amer. Ass. of Cost Engineers, Univ. of Alabama, Ala.).
GARFIELD, E. and SHER, I. H. (1967), *Diagonal Display — A New Technique for Graphic Representation of Complex Topological Networks*. Institute for Scientific Information, Philadelphia.
GARFIELD, E., SHER, I. H. and TORPIE, R. G. (1964), *The Use of Citation Data in Writing the History of Science*. Institute for Scientific Information, Philadelphia.
GOLAY, M. J. E. (1953), Referenced-author lists. *Physics Today*, 6/1, 20.
GORE, D. (1966), The mismanagement of college libraries. *Bull. Amer. Assoc. Univ. Prof.*, 52/1, 46.
LEDERBERG, J. (1959), Private communication, May 9.
LEDERBERG, J. (1963), Molecular biology, eugenics, and euphenics. *Nature*, 198, 428.
LUHN, H. P. (1959), *Keyword-In-Context Index for Techical Literature (KWIC)*. International Business Machines Corp., Advanced Development Division, Yorktown Heights, N.Y.
MERTON, R. K. (1969), Behavior patterns of scientists. *Amer. Scholar*, 38/2, 197.
NEWELL, A. (1962), Private communication, February 27.
PRICE, D. J. DE S. (1965), Networks of scientific papers. *Science*, 149/3683, 510.
SCHAUMBURG, H. H., BYCK, R., AMBOS, M., LEAVITT, N. R., MARMOREK, L. and WOLSCHINA, S. B. (1968), Sin cib-syn: accent on glutamate. *New Engl. J. Med.*, 279/2, 105.
SCHAUMBURG, H. H., BYCK, R., GERSTL, R. and MASHMAN, J. H. (1969), Monosodium L-glutamate: its pharmacology and role in the Chinese restaurant syndrome. *Science*, 163, 826.
SMITH, J. F. (1964), Systematic serendipity. *Chem. Engng News*, 42/35, 55.
SPENCER, C. C. (1967), Subject searching with *Science Citation Index*: preparation of a drug bibliography using *Chemical Abstracts, Index Medicus*, and *Science Citation Index* 1961 and 1964. *Amer. Doc.*, 18/2, 87.
STEINBACH, H. B. (1964), The quest for certainty: Science Citation Index. *Science*, 145/3628, 142.
STONEHILL, H. I. (1965), *Science Citation Index*: information retrieval by propinquity. *Chem. and Indus.*, 10, 416.

"Current Comments"

Citation Indexing and the Sociology of Science

Click for : "Citation Indexing, historio-bibliography, and the sociology of science", p158 Essays, Vol:1

April 14, 1971

The delay in publication of scientific papers is a constant source of frustration for their authors. Perhaps no segment of the literature is subjected to greater publication delays than that which eventually appears in the bound volumes emanating from international meetings and symposia. In May 1969, I presented a paper[1] which brought together much of my theoretical and practical work on the subject of indexing. During the twenty months it took to publish that work I was not scooped, as so often happens these days, but a number of developments did take place which made it obsolete without an appropriate supplement. I tried to rectify the situation by publishing a short paper in *Nature*[2] which has been reprinted in *Current Currents*® [3]. Indeed, the subject has been anonymously editorialized (a regrettable practice) in *Nature*[4] and amplified by Cawkell[5]. I in turn have responded to the challenge with a letter[6] whose publication has been delayed by the British postal strike.

Of even greater significance, insofar as the topic concerns science policy studies or the sociology of science, there have appeared in sociology journals not usually seen by most *CC*® readers a series of studies which provides almost incontrovertible support for the claim that citation analysis can be correlated quite well with other more subjective methods of analysis. Perhaps the most striking study to follow up the work of Cole[7] is that of Hagstrom[8], who has obtained an amazingly high correlation with Cartter[9] in evaluating graduate academic departments. These and several other relevant papers[10-15] are listed below and will be discussed in the future.

Following this editorial, the paper I originally read in Amsterdam is reprinted in its entirety. Quite frankly, this is the most economical method of getting it into the hands of those who have expressed interest in the topic. As is our custom, reprints are available.

1. Garfield, E. "Citation Indexing, Historio-Bibliography, and the Sociology of Science." in *Proceedings of the Third International Congress of Medical Librarianship, Amsterdam, 5-9 May, 1969*, ed. by K.E. Davis & W.D. Sweeney (Excerpta Medica, Amsterdam, 1970) pp. 187-204.
2. Garfield, E. Citation indexing for studying science. *Nature* 227:669-671, 1970.
3. *Current Contents/Life Sciences* 13(46):45-51, November 18, 1970.
4. Anonymous. More games with numbers. [An editorial in] *Nature* 228(5273):698-699, 1970.
5. Cawkell, A.E. *Science Citation Index*. [A letter to the editor of] *Nature* 228(5273): 789-790, 1970.

6. Garfield, E. Where the action was, is and will be. *Nature*, in press.

7. Cole, S. & Cole, J.R. Visibility and the structural bases of awareness of scientific research. *American Sociological Review* 33(3):397, 1968.

8. Hagstrom, W.O. "Inputs, Outputs, and the Prestige of American University Science Departments." Paper presented at the Annual Meeting of the American Association for the Advancement of Science, Chicago, December 28, 1970.

9. Cartter, A.M. *An Assessment of Quality in Graduate Education.* (American Council on Education, Washington, D.C., 1966).

10. Cole, S. & Cole, J.R. Scientific output and recognition; a study in the operation of the reward system in science. *American Sociological Review* 32(3):377-390, 1967.

11. Bayer, A.E. & Folger, J. Some correlates of a citation measure of productivity in science. *Sociology of Education* 39(4):382-390, 1966

12. MacRae, D. Jr. Growth and decay curves in scientific citations. *American Sociological Review* 34(5):631-635, 1969.

13. Parker, E.B., Paisley, W.J. & Garrett, R. "Bibliographic Citations as Unobtrusive Measures of Scientific Communication." (Stanford University Institute for Communications Research, Stanford, 1967, 125 pp.)

14. Price, D.J. deS. Is technology historically independent of science? A study in statistical historiography. *Technol. Culture* 6(4):553-568, 1965.

15. Whitley, R.D. Communication nets in science; status and citation patterns in animal physiology, *Sociological Review* 17(2):219-233, 1969.

Essays of an Information Scientist, Vol:1, p.133-138, 1962-73
Current Contents, #33, November 18, 1970 Reprinted From: Nature, 227: p.669-71,
Reprinted from *Nature* 227:669-71, 1970

Citation Indexing for Studying Science

Click for : Introduction on p. 132, Essays Vol:1

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 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, 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.[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

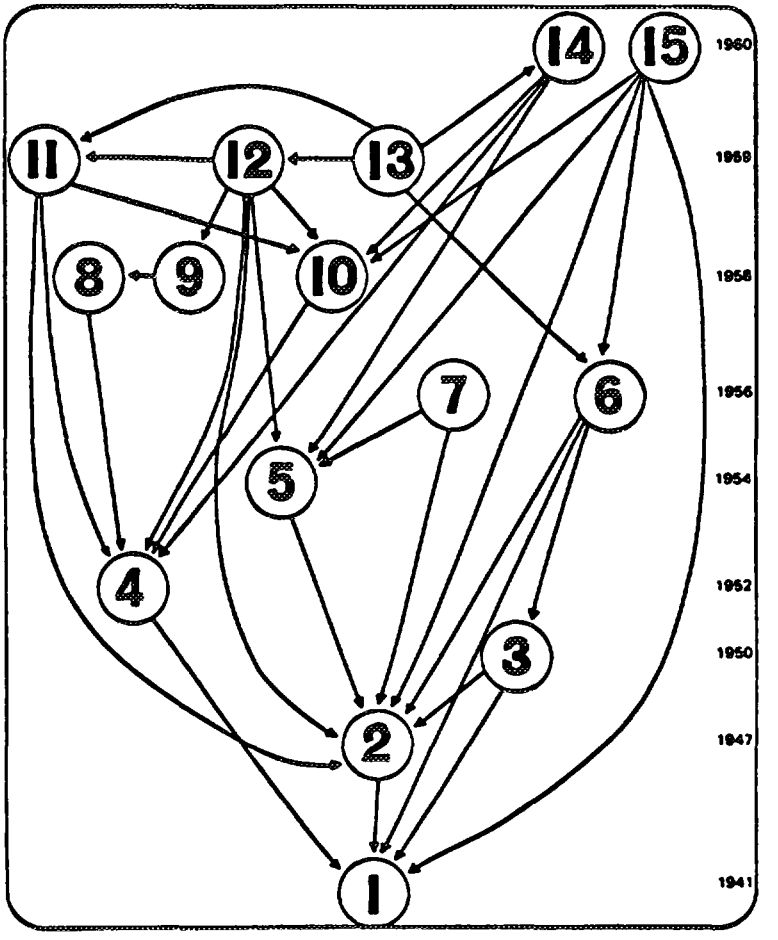

Figure 1. Citation network of fifteen articles on nucleic acids.

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 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

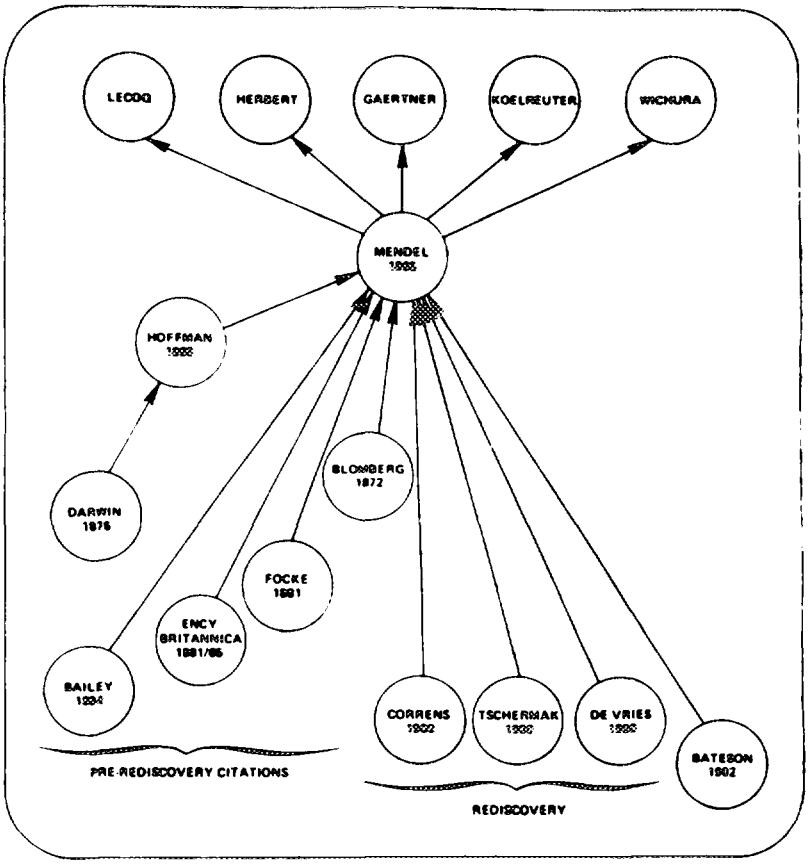

Figure 2. Citation network showing citations to Mendel before alleged "rediscovery". This was discussed by Zirkle in ref. 11a.

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.

Rank	Times Cited	Name	Rank	Times Cited	Name	Rank	Times 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	Eliel 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 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.

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.