*Social Science Information Studies* (1981), *1*, (263–280)


# BOOK REVIEWS

William Goffman and Kenneth S. Warren.
*Scientific information systems and the principle of selectivity.* New York: Praeger, 1980. vii, 189 pp. £13.00

The literature of science has been, and still is, growing. Much of it is of low quality, and in a given field the best articles tend to be concentrated in a small number of journals, nearly all of them in English. It would be useful for researchers, and also for libraries, if the more significant authors, articles and journals could be identified. Methods of filtering include an analysis of review articles, analysis of citations to selected papers and authors, and assessment by other workers in the field.

The above brief summary of the book under review will come as a surprise to few. That the text occupies 185 pages is partly due to the fact that some 55 of them are occupied by figures (most of which could have been reduced to half the size) and some 30 by tables. Further allowance must be made for repetition (the sensations of *déjà vu* experienced when reading this book are not at all mysterious). If the authors are as concerned as they say with saving the time of users, they could have considerably reduced the length of their book.

It is tempting to suggest that they might have reduced it still further by avoiding ground covered by previously published articles by Goffman (with or without colleagues), which have expounded over a period of 16 years what I take to be the core of the book—the exposition of the mathematical model representing the dissemination of information on the analogy of a 'four-stage epidemiological process'. The model has been elaborated from earlier formulations and there is in any case nothing wrong with synthesizing in a book the results of previously published research, but it does seem a little odd that reference is made to only one of at least twelve previous articles that are concerned not only with the model but with much of the remaining contents of the book.[1]

The book consists of a rather superficial, and sometimes questionable, introductory chapter; an exposition of the mathematical model of scientific communication; a detailed 'epidemiological analysis' of comprehensive bibliographies of three subjects (schistosomiasis, mast cells, and symbolic logic), which served as a test for the model; a qualitative assessment, by significant contributors to the field, of the literature of schistosomiasis; proposals for 'quality filtering'; and a final chapter 'Towards a rational policy for improvement of scientific communication'.

The analogy between literature dissemination and the course of epidemics is interesting. The model stood up well to the tests, though its full validity could be established only by testing against a larger number and wider range of literature. The bibliometric analyses are of interest in their own right, since few such analyses have been made of comprehensive bibliographies; their interest is by no means confined to the 'epidemic curves' that are traced. The qualitative assessment also yielded interesting results, although the reporting leaves something to be desired (see below). If the findings largely confirm what was already known, they do so from a different approach and on the basis of an impressive collection and analysis of data.

Two fundamental questions need to be asked. The first is whether the epidemiological analogy, if it is valid, is more valid than other possible analogies, for example the spread and decline of plant or animal species or of human cultures. The essence of them all is the process of spreading and then shrinking or dividing, a common enough process in biology and human life; in the case of research, this process is reflected in the published literature.

The second question is whether the model, or any model of this kind, is of any practical use whatever. The authors state their own belief that, while a drastic interference with the 'ecology of scientific communication' could have serious effects, dissemination should be somehow controlled, in the interests of the user, and that some kind of quality filtering is therefore desirable. In fact, they do not propose that their model can or should be used directly for the purpose. Rather, in their penultimate chapter they put forward several practical suggestions, nearly all of which could have been made without any of the modelling, bibliometric analysis or qualitative survey that constitute the bulk of the book.

Their suggestions fall into three categories. The first set, which constitutes an 'automatic procedure for quality filtering a scientific literature', involves in essence the identification of all review articles on a subject, an analysis of all references in them to produce distributions of cited authors, papers and journals, and the computation of 'quality numbers' for each. The weaknesses of this are obvious. There may be no recent review articles relating to the subject in question, some reviews aim to be comprehensive in their bibliographies (though critical in their text), and at best the results are retrospective rather than prospective. Good new papers will be published, and good new authors emerge. Some stability may be expected in the list of core journals, but these will probably already be known. The authors admit that 'the current literature will be missed by the reviews', but comment rather feebly that 'since quality is a function of time, the most current literature is too new to be accurately assessed for quality'. Apart from the fact that it is surely the *recognition* of quality that is a function of time, a filtering system that produces a thin pure stream of old water and leaves users to drown in floods of current water does not seem likely to be of much use.

The second set of suggestions relates to information retrieval, and relies on following up citations to known relevant articles and on co-citation links between articles. This method works well, but is hardly novel. The third set of suggestions is more ambitious; it is intended to aid science funding policy decisions by identifying 'authors whose quality numbers are increasing'. Whether the somewhat complex procedure, which involves the construction of 'epidemic curves', produces significantly better results than present methods seems doubtful.

The need for quality filtering has been quite frequently stated, but it begs some questions. Some researchers do not want their input filtered, whether because they do not trust the assessment of others (however 'objective' this assessment appears), because they recognize that even a poor paper can contain a small genuine experimental contribution to knowledge in their field, or because their thought processes may be more stimulated by poor papers, with seemingly crazy hypotheses, than by 'good' ones. (I found the most questionable parts of the present book the most stimulating because they forced me to think precisely what was wrong with them and how it might be put right.) Secondly, most of the 'quality' literature is produced in a very few countries. To eliminate the rest would leave developed countries with no training grounds for new authors and developing countries with no journals at all. This would confirm and extend the gap between rich and poor nations, north and south. Perhaps 'quality' scientists

should deliberately place good papers in the journals of developing countries to help them gradually reach a level that may otherwise never be attainable, with disastrous results for the economic as well as scientific and technological development of those countries. Are we in fact looking at scientific communication in far too narrow and self-centred a context, with the immediate (western) user in mind rather than the ultimate benefit to mankind?

If selectivity *is* desirable, can this not better be done at the publication stage by a combination of full-text articles (recognized as 'quality' immediately) and synopses, with full text available on demand and the possibility of full publication if the demand is great enough? Even at present, a discriminating user can and does do his own selecting: he scans only a limited range of journals, follows up only a proportion of references retrieved from data-bases (confining them mostly to articles in English in western journals), and if necessary uses the SCI as a filtering or supplementary retrieval mechanism. It is only too easy to assume that a problem exists and then devise methods for solving it, and the book seems to have fallen into this trap. At the least, some recognition of wider issues and alternative approaches would have been welcome.

The bibliometric analyses can be plundered for a variety of purposes; it is a pity that such collections of data often remain only partially exploited, and although the concept of a bibliometric data bank is not a new one it might be worth resurrecting. While most of the findings are more or less predictable, some are not. For example, in the field of schistosomiasis the proportion of articles with more than one author increases with the quality of the articles. The number of papers per author seems to be constant over time, and the 'literature explosion' therefore seems to be due to an author explosion. However, the length of journal articles has decreased somewhat. *Tropical Diseases Bulletin* does a fairly good job of selecting quality articles in schistosomiasis for abstracting, though its performance is much better for earlier than for more recent articles. Of all the articles relating to schistosomiasis 44 per cent are in English, but 86 per cent of the core literature is in English.

The book thus contains a good deal of marginal interest. The same cannot be said for the quality of its writing. Comment was passed earlier on repetitiveness, but at least redundancy provides some check on consistency and accuracy, qualities that are by no means always present. Not only is the proof reading very poor in places, but the text itself is sometimes at fault. We read, for example, of one 'D. S. Urquart' (read D. J. Urquhart) who is said to have analysed loan applications in 1956 at the 'British Lending Library for Science and Technology in London'. The analyses in question were of requests made on the Science Museum Library (see Urquhart, 1958); the National Lending Library for Science and Technology came into being at Boston Spa in 1961, and the British Library Lending Division in 1973. The authors go on to say that although the study showed a high concentration on a relatively few journals, 'the acquisitions policy of the library was not revised. ... This incident also demonstrates the attitude of most librarians that the larger the collection, the better the library'. On the contrary, it reflects the practical and economic sense of setting up a central comprehensive collection of journals in support of the vast majority of libraries that are neither able nor willing to aim at large collections. The author's name is spelt wrongly in the reference as well as the text, which quotes some figures that are not in the article at all.

There are numerous other examples of carelessness that erode confidence in the book. On page 27, the authors quote from 'a recent book by A. R. Anderson

and N. D. Belknap', which turns out to be a journal article published in 1962. MEDLINE, at first correctly referred to, soon becomes and remains (and is indexed as) MEDLINES. In some places it seems clear that the text is wrong, but the correct wording is not certain. On page 149, for example, the suggested procedure for quality filtering, a critical part of the text, includes an instruction to 'feed all references to all reviews into a computing machine'. Should this be 'all references in *all* reviews' or 'all references to all papers in reviews'?

Worse, some of the experimental procedures are not precisely stated or clearly worded. For example, we are told how 'significant contributors' to the field of schistosomiasis were identified, but not how exactly they were then asked to select authors and articles. Were they asked to mention those they believed from personal knowledge to be of 'quality'? They could not have known more than a fraction of the nearly 10,000 articles published 1852–1962 with which they were (presumably) confronted: 345 papers were selected 6 or more times, and 471 5 or more times. It seems likely that it is the 345, not the 471, that were resubmitted for further examination, though this is not wholly clear from the text. It is even less clear why the resubmission took place at all, but it resulted in a total of 404 articles. At this point 'further careful pruning' took place; why, and how? This whole section is especially confusing because while the assessment was of 9,914 journal articles, the total bibliography consisted of 10,286 items, and this total is used for the calculation of percentages arising from the assessment. The issue is further complicated by the mention of a bibliography of schistosomiasis 1852–*1972*. In a work on scientific information, of all subjects, the utmost clarity and precision should be expected: this book falls far short of the required standards.

The authors are not only guilty of inaccuracy, carelessness and imprecision. They write as if few others had discussed scientific communication through the literature, even when there are published papers that are directly relevant to the topics under consideration. One striking example is the discussion of measures of informational overlap between articles. It seems unbelievable that the authors did not know of Kessler's seminal paper on bibliographic coupling (1963a and b) or Small's (1973, 1977) later writing on co-citation, but scarcely less unbelievable that if they did they should not make any mention of either writer. Astonishingly little reference is made to use of the *Science Citation Index* data-base for retrieval and filtering purposes.

The index is poor. It has no entries under names of authors cited, and lacks other useful entries. For example, the performance of *Tropical Diseases Bulletin* and the proportions of the literature in different languages were mentioned above. Finding the relevant parts of the text would be difficult, because there are no entries under *Tropical Diseases Bulletin*, 'language', or 'English', 'German' etc. On the other hand, some entries seem useless: who would look up 'poor interpretation of data' or 'Librarians: attitude of'?

Whether it is worth spending £13 on this book depends largely on whether one is prepared to pay this much for access to the mass of useful bibliometric data it contains.

Maurice B. Line

## NOTE AND REFERENCES

[1] Previous articles by Goffman include:

GOFFMAN, W. and NEWILL, V. A. (1964). Generalization of epidemic theory—an application to the transmission of ideas. *Nature, 204*, 225–228.

GOFFMAN, W. (1966a). Stability of epidemic processes, *Nature*, *210*, 786–787.
GOFFMAN, W. (1966b). Mathematical approach to the spread of scientific ideas—the history of mast cell research. *Nature*, *212*, 449–452.
GOFFMAN, W. and NEWILL, V. A. (1967). Communication and epidemic processes. *Proceedings of the Royal Society, Series A*, *298*, 316–334.
GOFFMAN, W. (1968). An indirect method of information retrieval. *Information Storage and Retrieval*, *4*, 361–373.
GOFFMAN, W. and WARREN, K. S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature*, *221*, 1205–1207.
GOFFMAN, W. (1971). A mathematical method for analyzing the growth of a scientific discipline. *Journal of the Association for Computing Machinery*, *18*, 173–185.
GOFFMAN, W. and HARMON, G. (1971). Mathematical approach to prediction of scientific discovery. *Nature*, *229*, 103.
GOFFMAN, W. and SARACEVIC, T. (1977). Structure and behavior of subject literatures as the base for forecasting in scientific communication. *International Forum on Information and Documentation*, *2*, 17–19.
KESSLER, M. M. (1963a). Bibliographic coupling between scientific papers. *American Documentation*, *14*, 10–25.
KESSLER, M. M. (1963b). Bibliographic coupling extended in time: ten case histories. *Information Storage and Retrieval*, *1*, 169–187.
SMALL, H. G. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*, 265–269.
SMALL, H. G. (1977). A co-citation model of a scientific speciality: a longitudinal study of collagen research. *Social Studies of Science*, *7*, 139–166.
URQUHART, D. J. (1958). Use of scientific periodicals. *In: International Conference on Scientific Information, 1956. Proceedings.* Washington, D.C., National Academy of Science, 1958, pp. 288–300.

Samuel Long (Editor).
*The Universal Reference System: Annual Supplement, 1979*. New York: IFI/Plenum, 1980. 3v. (*Political Science, Government, and Public Policy Series.*) ISBN 0 306 69029 2 $350.00.

Prodigious amounts of professional dedication, time and scarce money are devoted currently to the world-wide business of producing and publishing, in ever increasing numbers, bibliographical tools of various descriptions. It would be comforting to believe that such bibliographical artefacts constituted serious and considered contributions likely to facilitate useful intellectual connections through the complex processes of the formal communication system. Chastening experience suggests that such a belief places an insupportable strain on bibliographical credulity. This, the 13th *Annual Supplement* to the *Universal Reference System* does little to sustain one's bibliographical faith. It is of the modern class of bibliographical productions which appear to operate on the assumption that a bibliographical problem will have to succumb if sufficient bibliographical items are thrown at it. A glib statement? Then let us examine the case in greater detail.

What have we the right to expect of a bibliography, or, in this instance, a bibliographical supplement? Accuracy of detail must be regarded as essential;