

Foreword

by
Derek J. de Solla Price

I was inoculated with Citation Fever shortly after coming to Yale and delivering the lectures that became *Science Since Babylon* (1961). In those days, the early 1960s, the National Science Foundation had a Science Information Council, which was supposed to advise and consult on policy questions concerned with funding the scientific literature. One may question how much impact we had on policy, but there is no doubt that the Council was one of the best information and education channels to which I have ever had access.

One memorable day that unforgettable character Gene Garfield appeared before the Council with a request to support the printing and distribution of the first experimental *Science Citation Index*, the data for which had been created to compile the *Genetics Citation Index*, funded by NIH. Unfortunately, NSF refused the request. Notwithstanding the refusal, I personally was immediately struck by the realization that citation links represented a radically new kind of data with far-reaching potential. Though we couldn't predict with absolute certainty how much a citation index might be used, or even to what purpose, it seemed clear to me that such an index must be developed. It also seemed clear to me that such an index would have a good chance of becoming a commercial success, instead of becoming a permanent burden on the federal budget; though a new immigrant to the land of federal fiscal matters, I was able to recognize that prospect as being nearly unique.

From that day to the present, which offers me the privilege of contributing this foreword in honor of Gene and his venture, I have found megavitamins for my intellectual diet on the cutting-room floor of ISI's computer room. Bit by bit we have begun to understand how citations work, and in the course of this, there has emerged a new sort of statistical sociology of science that has thrown light on many aspects of the authorship, refereeing, and publication of scientific research papers. The Society for Social Studies of Science (known colloquially as the 4S Group) now has an annual meeting devoted to this new method of understanding science that has grown, almost as an accidental by-product, from the indexing technology developed by the Institute for Scientific Information. Our initial intuitive perceptions have turned out to be correct.

One thing that we failed to perceive in those early years, however, was the open-ended character of the computer revolution. It was reasonably supposed at that time that the computer would soon have the capacity to do various large memory jobs, such as computer indexing, and that it would gradually do them faster and faster as well as cheaper and cheaper. What was not perceived was that the rate and magnitude of hardware advances would be such that it would become possible to do classes of jobs one year that had not even been conceptualized the previous year. It is rather like the early days of some other technologies; it would have been difficult to perceive that the typewriter would generate a new sociology of office workers, or that the automobile would invent suburbs.

What I see now with the new understanding of how citations and papers work, and with a more realistic appreciation for the technological potential of computers, is that one may use citations as a way of organizing and filing the scientific literature as an omniscient and perfectly-read scholar in the field would do. All this derives from the discovery by Henry Small and Belver Griffith that co-citation mapping can order papers in clusters that map on a two-dimensional plane. It may seem a rather abstruse finding, but I believe it to be revolutionary in its implications.

The finding suggests that there is some type of natural order in science crying out to be recognized and diagnosed. Our method of indexing papers by descriptors or other terms is almost certainly at variance with this natural order. If we can successfully define the natural order, we will have created a sort of giant atlas of the corpus of scientific papers that can be maintained in real time for classifying and monitoring developments as they occur.

I believe it will soon (in five years or so) be possible to display pages from such an atlas, showing not only the natural place of each new paper in "knowledge space" but also giving the degree and nature of the activity resulting from each contribution. As in air-traffic-controller displays, each element in the atlas display can be tagged—in this case with the nation, institution, person, and granting program associated with each paper.

The new mapping theorem has both theoretical and practical implications. On the side of pure speculative theory I suggest that *information is measured, as we well know from Shannon theory, by the order it produces out of disorder. But order of what? The answer seems to be that each piece of information has value insofar as it relates to the order of other information, and that what we see in mapping is this basic order.* That order has been invisible till now, because the structure of previous information systems has constrained information to a single dimension. When we let information spread out into a second dimension, as on a plane, the order becomes visible. I, for one, find that very challenging and provocative.

At the practical level I feel we shall be able to use the mapping theorem for indexing. At an even more elementary level, it suggests a useful condensation of printed citation indexes. We already know from cumulative-advantage theory that knowledge grows almost entirely from the small core of highly-active papers that are more than minimally cited. The co-citation mapping work demonstrates the fundamental nature of that core and raises the question of whether it is worthwhile to include anything more than that in printed citation

indexes, when we know that the rest has a very low utility and creates functional and economic problems. Mightn't we produce smaller, easier to handle, and less expensive printed indexes by restricting entry to those papers cited more than once or, in annual and larger cumulations, perhaps even to those cited by three different source papers. Besides reducing the bulk and cost of the printed volumes, such a strategy would automatically exclude all the very annoying garbage that is introduced by mis-citations, bibliographic and transcription errors, and the half of all references that are scattered around as random noise. The full citation file could be maintained on-line for those unwilling to trade off unnecessary completeness for a much higher degree of relevance.

This volume of essays contains one item that touches on this very point of condensation. The one on *Project Keysave* describes a technique that has proved very useful and economical at ISI. A select file of the more highly cited papers is combed through automatically as each new reference is keyed into the system in abbreviated form. For many journals 90 percent of all references can be "recognized" at this early stage of inputting, eliminating the need for additional keying. One idea suggested by this methodology is that the contents of this file of constantly-used references would make a valuable personal possession for most publishing scientists. I suppose that if each good old standby paper in the file were issued with a social security number, future authors could cite that and avoid all the usual misidentifications.

Amongst other essays reprinted in this volume, I would like to draw the reader's special attention to those dealing with citation studies, in particular, the one on the study of the parochial character (or otherwise) of French science, "Le Nouveau Défi Américain," which drew scandalized protests from the French undoubtedly because of the special political significance they have always attached to the international civilizing mission of the French language. Similar studies of other nations, areas, and languages make an impressive and policy-provoking set of points.

I think it will be necessary one day to organize these insights in a way that makes them useful at the national-policy-making level. What should the government of Brazil do about its scientific journals, for example, in the light of this citation data? Should they publish their best work in the Portuguese language? Almost certainly not. Should they continue to claim, as they do in their national bibliography, that they publish some 3000 scientific and technical journals? It is hardly useful when one cannot find citation evidence that more than a handful of these are ever used or cited by anyone in their own country.

These essays document the peculiarly strategic role Gene Garfield and ISI play in using the computer to deepen our understanding of scientific information and the natural order of scientific knowledge. They are not just the best **ballgame** in town, but the only one in the sense that they are covering all of science rather than some particular fragment of it, and in the sense that they have developed, and are evolving, a perspective of the whole that is unique to the computer manipulation of information. My guess, for what it is worth, is that the continuing development of computerized information will carry their effort forward for many decades to come, and the provocative observations and useful insights collected in this volume will be followed by many more.



Derek J. de Solla Price
Department of History of Science and Medicine
Yale University
Box 2036
Yale Station
New Haven, Connecticut 06520