Essays of an Information Scientist, Vol:2, p.26-31, 1974-76 Current Contents, #7, p.5-6, February 13, 1974

"Current Comments"

ISI Is Studying the Structure of Science Through Co-Citation Analysis

February 13, 1974 Number 7

What are the patterns of scientific research and how do they shift and change? This problem in the sociology of science is attracting the attention of increasing numbers of information scientists and other researchers, especially in the social sciences. A bright young historian of science, Dr. Henry G. Small, research associate at the Institute for Scientific Information has a deep interest in this problem. Recently he made a significant contribution to the methodology of this subject in an article reprinted below.[1] Such studies, which look at science from various vantage points, have significant implications for a number of fields. In information science, for example, the structure of science has a direct impact on the design of classification systems and the ability to cope with the changes required in them. Sociologists are concerned with the structure of science in terms of patterns of informal communication, group structure (invisible colleges), etc. In science policy studies there is a need to know more about 'where the action is,' who is involved, and where it is headed. Organizations like NSF and Congress need to understand the links between various specialties. Finally, historians and philosophers of science are, more and more, talking about 'paradigms', 'clusters', and 'communities', that is, the social and collective aspects of scientific activity.[2]

In much of this work, there is increasing recognition that citation data provide a significant if not one of the best means for studying the structure of science, protestation not withstanding.[3-4] The groundwork for much of these current efforts was pioneered by Price.[5] My own work has been variously described as 'science mapping' or the 'geography of science.'[6] One goal of such research is the spatial display or representation of the entire mosaic of specialties in the natural sciences.

Much of the current work uses as its data base the magnetic tape version of the multidisciplinary *Science Citation Index®* (*SCI®*). By applying known methodologies of numerical taxonomy with clever computer techniques, clustering studies of great interest can be done. The *SCI* data base will undoubtedly increasingly establish itself as a vital data base for the study of science.

The article reprinted here is the first step in a science mapping project currently underway at ISI® under a grant from the National Science Foundation. In the article, a new method is introduced for establishing the linkage or association between two frequently cited papers. The new measure is called "co-citation." Co-citation is to be compared with a similar but distinctly different measure called "bibliographic coupling" introduced by Kessler in 1958. (In a formal treatment of citation networks, co-citation emerges as the mirror image of bibliographic

coupling.) Unlike bibliographic coupling, co-citation can be readily determined from a printed citation index such as the *SCI*. This is done by counting the number of identical citing items listed under two selected cited items. Two cited publications are "co-cited" when at least one paper cites both. The more co-citation one observes, the more closely related the cited items are. A number of procedures can be used to "normalize" this measure of relationship (that is, eliminate the effects of sheer size), and to convert what is a measure of proximity or closeness to a measure of "distance." This latter step enables one to talk about the "space" occupied by science and its characteristics (for example, its dimensionality) and leads directly to the notion of mapping.

An example of a cluster in particle physics is given in the article. (This is not surprising considering Small's former experience at the American Institute of Physics.) The co-citation links among the papers were determined by manually counting them in the *SCI*. This procedure has now been fully automated.

The research team at ISI has been able to generate clusters in nearly all active specialties in science from cancer research to plate tectonics. When fully developed, this clustering system promises to provide some new capabilities. For example, we envision the automatic creation of classification schemes, and the *automatic* classification of journal articles. Most readers of *Current Contents*® will not appreciate that there is a large literature on "automatic classification" which describes many such methods, most of them anything but automatic.

For science policy studies a system will be at hand for the continual monitoring of scientific developments and break-throughs and supplying of supporting demographic data (individuals and institutions). These are by no means limited to "most-cited" papers or authors.

Having found a way to define a structure for science we can now tackle the problem of structural change. Since clusters can be identified on a quarterly or annual basis, it becomes possible to relate clusters obtained in one period with those obtained in the next. In this perspective, a network like the one given in the reprint is a snapshot at one point in time of a dynamic system which is moving in the "space" which defines science. Clearly, we must study a succession of snapshots before we can hope to understand the mechanisms and processes of change. This work is just beginning and will take a concerted effort on the part of many researchers. The ultimate question–whether we can anticipate the future of science–will depend in large part on our success in finding recurrent patterns of change in the past.

1. Small, H., Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Amer. Soc. Inform. Sci.* 24(4):265-9, 1973.
2. Cole, J.R., & Cole, S. The Ortega hypothesis. *Science* 178:368, 1972.
3. Janke, N.C. Journal evaluation. *Science* 182:1196-7, 1973. See rebuttal: Garfield, E. Journal evaluation, *Science* 182;1197-8, 1973.
4. Goudsmit, S.A. Citation analysis. *Science* 183:28, 1974. See rebuttal: Cole, J.R. & Cole, S. Citation analysis. *Science* 183:32-3, 1974.
5. Price, D.J.D. Networks of scientific papers. *Science* 149:510-15, 1965.
6. Garfield, E., Sher, I.H. & Torpie, R. J. *The Use of Citation Data for Writing the History of Science.* Philadelphia: Institute for Scientific Information, 1964, 86 pp.