This research was an attempt to develop a quantitative measure of scientists' active and changing perceptions of the associations of scientific ideas. After initial work using title word co-occurrences, the focus was changed to an analysis of citations, which proved to provide an easier and more accurate tool. (The *SSCI*® and the *SCI*® indicate that this paper has been cited in more than 145 publications, making it the most-cited paper published in this journal.]

Cogitations on Co-Citations

Henry Small
Institute for Scientific Info mation
3501 Market Street
Philadelphia, PA 19104

This paper was the product of my first few months at the institute for Scientific Information® (ISI®), under the guidance of Eugene Garfield and Morton Maiin. It is also my first paper, and, in a sense, I am still writing it

I had some preliminary notions of what I wanted to do at ISI from a previous project at the American Institute of Physics. As a historian of science, I was looking for a way to systematiccalry use the published physics literature to map out the early history of nuclor physics.[1] under the influence of Thomas S. Kuhn,[2] I hoped that by using data from scientific texts, such as key words, references, or classification headings, I could delineate a paradigm in the field, and then observe its evolution. From my predilections in the history of science, I was seeking the best possible quantitative measure of the association of scientific ideas. Unlike M.M. Kessler's bibliographic coupling,[3] I sought a measure that reflected scientists' active and changing perceptions of these associations.

My first inclination was to use title word co-occurrences as, for example, embodied in ISI's *Permuterm® Subject Index,* and now dubbed co-word analysis by French sociologists.[4] But when I realized that ISI was the only organization in the world that maintained a citation index database, I quickly switched from key words to citations. It was a fortuitous choice because it turned out that cited references were not only more precise and had fewer linguistic ambiguities than title words, they were an ideal bridge ~~between~~ the cognitive world of scientific ideas and the social world of scientists. The term "co-citation" seemed a natural way to describe the measure.

In gathering my data, I worked with the printed *Science Citation Index®,* scanning pairs of columns for common entries. The subject relationship among highly co-cited papers was surprisingly dose. Garfield took a keen interest in my initial draft, suggesting many editorial changes. One comment of his I recall is that my use of the word "new" in the title would look odd from a perspective of 20 years, which at the time seemed too far in the future to be of concern. Another comment came from a researcher in Pittsburgh who said that I had not used a property algorithmlc procedure in producing my co-citation cluster. I resolved to learn cluster analysis and the necessary computer skills to automate the procedure. At this time I was also fortunate to meet Belver Griffith, of Drexel University, who not only made useful comments on the paper but became a dose collaborator on subsequent developments of the methodology.

As is often the case, we later find numerous co-discoverers or precursors for what we thought were our ideas. Shortly after the paper was published, Tony Cawkell, ISI's director of research, brought to my attention a paper in Russian by an information scientist named Irene Marshakova.[5] She had used essentially the same method to map the structure of a research field. Finally, after many years, I had the pleasure of meeting Irene when she visited ISI. She was still working by hand, never having had the benefit of computer technology. Also, in reading some discussions from the early years of information science, I came across a suggestion by the linguist Y. Bar-Hillel, that someone ought to try a "co-quotation" method.[6] I wrote Bar-Hillel in Israel to see if he had developed the idea further, but learned that he had suffered a severe stroke, from which he later died. Don R* Swanson recently noted this omission from my reference list.[7]

The reason this paper has been cited so frequently is that co-citation has proved to be a useful tool in the analysis of scientific informstion. Of course, the PR that I have received at ISI over the years hasn't hurt either.

1. **Small H.** Bibliometric indicators of the development of nuclear physics: *The Physical Review,* 1927-1934. *American Institute of Physics,* 1971. (Unpublished report.)
2. **Kuhn** T S. 77κ *structure of scientific revolutions.* Chicago, IL: University of Chicago Press, 1970. (Cited 6,540 times.)
3. **Rosier M M.** Bibliographic coupling between scientific papers. *Amer. Doc.* 14:10-25,1963. (Cited 75 times.)
4. **Callon M, Law J & Rip** A, eds. *Mapping the dynamics of science and technology.* London: Macmillan Press, 1986. (Cited 45 times.)
5. **Marshakova** I V. System of document connections based on references. *Nauchn—Tekhn. Inform. Ser. 2 SSR* 1973(6):3-8.
6. **Bar-Hillel** Y. *Proceedings of the International Conference on Scientific Information.* Washington, DC: National Academy of Sciences, 1959. p. 1,408.
7. **Swanson D R.** Historical note: information retrieval and the future of an illusion. *J. Amer. Soc. Inform. Sci.* 39:92-8. 1988.
Received February 7, 1992