Reprinted in Essays of an Information Scientist, Vol:6, p.580, 1983 from
Toward a Metric of Science: The Advent of Science Indicators, Eds. Yehuda Elkana, Joshua Lederberg, Robert K. Merton, Arnold Thackray and Harriet Zuckerman, John Wiley & Sons, NY, 1978

8

Citation Data as Science Indicators

Eugene Garfield
Morton V. Malin
Henry Small

Attempts to appraise the condition of science as intellectual activity or social institution have involved data compilation—reports on amounts of money spent on scientific research, magnitudes of scientific manpower, number of students enrolled as science majors at universities, and number of scientific papers produced or number of patents issued. For a variety of reasons, these all fail to indicate the "condition" of science. Perhaps the problem is that such data are compiled and presented without regard to a *specific* set of questions or set of hypotheses; thus a coherent framework for estimating the social or intellectual condition of science is missing. *Science Indicators 1972 (SI-72)* is an improvement over mere data compilation, and it should be applauded as a step, however preliminary and tentative, in the right direction. The purpose of this paper is to suggest further indicators relevant for measuring scientific activity, in the hope that this will lead to a better estimate of the condition of science.

At the Institute for Scientific Information (ISI), we operate on the fundamental as-

We thank Professor Belver Griffith of Drexel University, who participated in the mapping-science project and has contributed to the development of these ideas. Research was supported by a National Science Foundation grant to the Institute for Scientific Information.

sumption that citation data can be used as indicators of present, past, and perhaps future activity in science (1). The validity of this assumption must of course be tested; but at present standards against which the validity of citation analysis can be measured do not exist. All that can be done, and perhaps all we can expect to do, is to compare the results of different methodologies and attempt to find significant correlations between them. Thus, for example, Hagstrom (2) found a significant correlation between citation analysis and subjective peer judgment in the Cartter report on quality of science graduate departments of American universities.

CITATION ANALYSIS

Citation analysis is a bibliometric method that uses reference citations found in scientific papers as the primary analytical tool. Bibliometrics can be defined as the quantification of bibliographic information for use in analysis (3). The literature of science lends itself to quantification because each source article, report, note, book, and so on contains such bibliographic elements as authors' names, addresses (country, state, city, institution, department), titles (words and phrases), journal titles, place of publication, volume and page number, and date of publication. All are keys that when properly organized, provide a base for extracting analytical data. However, the added element of reference citations in scientific and scholarly literature is most significant for citation analysis.

Sources

The manipulation of all these bibliographic elements for sociological, historical, and other kinds of studies became practical on a large scale only when the computer entered the picture. Consider, for example, the difficulty of creating the *Science Citation Index (SCI)* data base without the aid of a computer. Approximately 3.5 million source items and 41 million citations have been input into this data base for the years 1961–1973. Approximately 400,000 source articles and book reviews, covering both current and past years, are added to the *SCI* data base each year.

Citation Data

Using the bibliographic elements extracted from the journal articles being processed, the information can be sorted in a variety of ways to provide data for studying a variety of questions and problems (see Kochen, also Cole et al., both this volume). Using citation data, Jonathan and Stephen Cole examined the phenomenon of social stratification among scientists, taking citation counts as a rough measure of peer recognition and of importance (interpreted as utility) of a scientific work (4). Derek Price has pioneered in the use of citation data in science policy studies and has made an important contribution to our understanding of citation networks (5). An early study of the use of citation data in historical research was done by Garfield, Sher and Torpie (6). Other studies, too numerous to mention here, are regularly reported in the *SCI* itself (7).

Science Mapping. Currently at ISI, citation data are being used to study the relative impact of scientific journals (8), the impact and quality of the research of individuals and institutions, and the specialty structure of science. The last-mentioned application, "science mapping," is of particular interest because the methodology extends the usefulness of citation data.

Certain presuppositions, both historical and sociological, underlie the idea of "mapping" science by identifying key papers and events through citation analysis. The basic unit of analysis in mapping is the highly cited document. The assumption is that these articles and books are markers for critical scientific ideas or events, taken in the broadest sense. This includes theoretical formulations, speculative hypotheses, experimental results, procedures or methods, and any combination of these. The fact that some documents have been highly cited within a specified time-period confers upon them a special status as providing important "ideas" in their respective areas or specialties. It should be possible to identify the corresponding cognitive components for each highly cited item either by examining the citing context or by querying the citing author.

The former was done for a sample of highly cited papers in chemistry (9). For each highly cited paper, a sample of citing papers was obtained. Each citing paper was examined to determine where in the paper the reference was made. The terminology used by the citing author in referring to the highly cited item was noted. In general, the more highly cited the item, the more uniform the terminology: terms such as *Hammond's postulate* or *atomic scattering factors* or *orbital symmetry rules* were invariably associated with certain works. These works had a clear conceptual identity for the specialists who cited them. Less highly cited papers, though not always achieving this level of eponymy, could nevertheless be associated with distinct conceptual entities. Our hypothesis is that most of, if not all, the scientific ideas that have been regarded as important or influential can be associated with one or more scientific works that are *at some time* highly cited. Sometimes recognition through citation frequency comes soon after publication, but a two- or three-year time-lag is the norm. This view does not rule out the possibility of a long delay before a work is recognized—perhaps exemplified by the Einstein 1905 paper on relativity (10)—nor does it rule out resurrection of older and perhaps long-forgotten works as the basis for new departures.

Definitions. The implication of the phrase *at some time* should be considered carefully. Papers containing important ideas will not necessarily continue to be highly cited for all time. Eventually, an idea or paper may become so widely known that citing its original version is unnecessary: Knowledge of this kind may be "tacit" in Polanyi's sense (11). Or a new paper will supersede the original one by reformulating the idea in more up-to-date terms; the newer paper then receives all the citations to the idea.

The hypothesis does not say that *all* highly cited papers contain important "ideas," in the narrow sense of the term. Clearly, papers can become highly cited for important methods, procedures, and data compilations; the fact that they are not "ideas" does not make them less important. The availability of a good table of nuclear masses, for example, is probably of critical importance to the advancement of nuclear physics. Finally, *important* should not be confused with *correct*, for an idea need not be correct to be important. This is evident in the recent polywater controversy (12). The high negative citation rates to some of the polywater papers is testimony to the fundamental importance of this substance if it could have been shown to exist.

In the preceding discussion we have left the term *important* undefined. Like *quality* of scientific work, *importance*, as perceived by a scientist, is undoubtedly a highly complex matter, even though scientists are constantly called on to make such judgments of the work of their peers. The intent here is not to define some absolute scale of importance as measured by citations, but to operationalize the notion so that we can compare citations and scientists' perceptions in terms of responses to questions such as: What are the most important advances in your specialty in the last five years? or How would you rate the quality of this paper on a scale from 1 to 10?

This view of the highly cited paper stresses its role as a marker of discovery or invention (see Kochen, this volume). That the number of such works is relatively small is demonstrable from statistics on highly cited papers. Even though the absolute number of items cited is high, only about 1% of the items cited in a year are cited 10 times or more. It follows that few such seminal works exist for any single specialty or research area; statistics on average cluster* size indicate there may be as few as five per area. The

*Typically, a cluster is a core of discovery papers surrounded by succeeding works built on the original discoveries (see next section).

usually high citation of these works over a period of years exhibits growth and decay characteristics that fit with intuitive feelings about the rate of obsolescence of knowledge or shifts in intellectual fashions. The set of highly cited works can even be seen as quite concretely representative of the paradigm for the specialty. Indeed, Kuhn (13) may have had this in mind.*

The normal turnover in highly cited works from year to year—that is, the appearance of new cited works and the disappearance of old ones—reflects, then, the rate of change of scientific conceptions about the world. However, change in documents need not be revolutionary, since one document may merely replace another by formulating the key idea in a more useful way. Generally, the more highly cited a paper is to begin with, the longer it continues to be highly cited. Methodologically important papers, like Lowry et al. on protein determination (14), are consistently the most highly cited papers in the *SCI* file. Kuhn's observation that such methodological tools are critical to the paradigm is certainly borne out by citation analysis.

If the Kuhnian paradigm is defined in terms of highly cited papers, the structure of the paradigm changes constantly—though slowly. A sample of highly cited works from a single source year identifies not only works that contain original discoveries but also a set of more recent derivative works built on the original discoveries that are more ephemeral. The totality of these founding and derivative works will represent the research front during the source year but not the entire history of a given subject. It is precisely this change, or turnover, in the set of highly cited papers which can become useful in studying the history of a specialty.

As an illustration, consider Bohr's well-known 1913 papers on atomic structure. The first of the papers (15) was only moderately cited (six times) in the 1973 *SCI*. But the failure to find Bohr's paper highly cited in 1973 means that the research front in atomic physics has moved on—presumably, Bohr was highly cited when his work set the framework for atomic theory in the early 1920s. The implication is not that Bohr's paper lacks "lasting value," but that in general, a paper or an idea may have critical importance for one historical period or environment and not for another. Citations are an "indicator" of a paper's importance or timeliness for a particular historical period; finding exactly what determines the timeliness of a paper or an idea for its particular conceptual and social context is a problem of key importance for the history of science.

*Because Kuhn is difficult to pin down, the reader may make his own interpretation from this passage in the original Preface: "Or again, if I am right that each scientific revolution alters the historical perspective of the community that experiences it, then that change of perspective should affect the structure of post-revolutionary textbooks and research publications. One such effect—a shift in the distribution of the technical literature cited in the footnotes to research reports—ought to be studied as a possible index to the occurrence of revolutions."

Highly cited papers are almost never isolated but tend to aggregate in small coherent groups. The tendency toward aggregation could be evidence for simultaneous discovery within an area, provided the papers had the same publication dates. Robert Merton has stressed the frequency with which "multiples" occur in science (16), and it appears that such a multiple will emerge in citation data as a cluster of highly cited documents. Successive as well as simultaneous discoveries will be related and grouped. Thus, the cluster will consist, typically, of a core of discovery papers (either simultaneous or successive) surrounded by several follow-up or derivative works, which have been built on the original discoveries. Any or all of these may be of theoretical, methodological or experimental import.

Co-citation and Bibliographic Coupling

Clearly, the way relations between the highly cited papers are derived is critically important for interpreting the significance of the groupings. Frequency of co-citation, which has been used to determine the relations among highly cited papers, is simply a count of the number of documents citing both of the highly cited documents in a specified time period (in this case, one year). Co-citation, therefore, reflects the association between highly cited papers as perceived by the current population of specialists who have themselves published papers. If, then, highly cited papers can be placed in some kind of correspondence (not necessarily one-to-one) with cognitive components (theories, experiments, methods, etc.), co-citation becomes a measure of cognitive association. In that case, the changes in such patterns of association from year to year can tell us something about the history of ideas.

At this point we might well pause and consider why some alternative methodologies were not employed. Considerable work in information science relates to the use of language associations, or term-term associations for defining related words and for clustering documents (17). To the extent that words can be associated with ideas, the results would be similar to those obtained by using co-citation. This could be done say, by counting the frequency of co-occurrence of words in titles of papers. But difficulties arise when the *SCI* is used: With a large multidisciplinary data base, words like *plasma* or *complex* which have different meanings in different fields, cause word associations to give rise to many false linkages between fields when such homographs are encountered. This weakness of natural language is not inherent in the reference citations in the form of document surrogates. The latter are virtually unique: No two of them can have the same author, journal, volume, page, and year. Hence, they are ideally suited for automatic manipulation.

Co-citation (18) should also be distinguished from another familiar coupling procedure, bibliographic coupling (19), to which it is related by a kind of mirror symmetry. In co-citation, earlier documents become linked because they are later cited together; in bibliographic coupling, later documents become linked because they cite the same earlier documents. The difference is that bibliographic coupling is an association intrinsic to the documents, the authors themselves having established it by citing one or more of the same works. Co-citation, in contrast, is conceived as a linkage extrinsic to the documents and one that is valid only so long as the community of specialists chooses to co-cite them. Thus, co-citation depends on the collective choices of a population of scientists who have published in the source year. Therefore, discovery papers, either simultaneous or successive, are highly co-cited—as well as being highly cited—and form a cohesive cluster. Further, the reason the groupings so derived correspond closely with what have been regarded as "specialties" is immediately visible, for only the "specialists" capable of understanding and utilizing the discoveries will cite and co-cite the discovery papers in their work. This identification of clusters with specialties is one of the firmest results to emerge from clustering studies using co-citation. Stated in different terms, discoveries or innovations are almost always specific to a problem area that is the domain of a specialty group. Notable exceptions exist: Certain methodologies (e.g., protein determinations) are applied in several specialties. Documents of this kind tend to span several specialties as a kind of methodological structure superimposed on a more fine-grained and intensely interactive specialty structure (20), like the role of function words in language analysis.

Clearly, bibliographic coupling and co-citation are closely related and derivable from a general treatment of pathways in a directed citation graph. Neither should be assumed to be implicitly superior as a measure of association or relatedness of documents. Each measure has its application. For example, since clusters obtained by using bibliographic coupling would presumably be no different from those obtained using co-citation, the only difference is operational—grouping together the citing papers in the former case and the cited papers in the latter. But once a cluster of citing papers is formed by bibliographic coupling, a list of the cited items the citing papers had in common can be generated; and likewise, once a cluster of cited papers is obtained by co-citation, the citing items can be retrieved. Since the procedures lead to the same end-result—namely, a clustering or classification of the literature—the choice of procedures is determined by the nature of the phenomenon to be investigated and the interpretation sought.

Social and Cognitive Structure of Science

From experiments on clustering using co-citation, the specialty appears to be a natural unit of structure and organization in science. Studies of the social structure of specialties, which have been pursued for some time now by sociologists, therefore appear most relevant and appropriate. These include studies of informal communications (21) and contacts among scientists and the so-called "invisible colleges" (22). To clarify the connection between citation studies and the social structure of science, the types of relations implied by the citation linkages must be considered. The three types of linkages are (a) direct citation, the citing of one document (scientist) by another, which is analogous to a sociometric choice—that is, highly cited papers have been "chosen" frequently and the authors of these works assume the role of leaders or "stars"; (b) co-citation, the equivalent of citing the "stars" together, which may or may not reflect the existence of informal contacts among the stars; and (c) bibliographic coupling, the choosing of one or more of the same stars another person has chosen. None of the bibliometric linkages require that social contacts lie behind them, but the existence of strong patterns of coupled documents (clusters) suggests that underlying social factors are at work. Just as a coherent body of knowledge about a fairly narrowly defined subject would be inconceivable without an underlying network of informal communications among specialists, a cluster of documents probably reflects an underlying social network. The extent to which these structures (documental and social) are congruent is not known.

The document networks derived through citation analysis are believed to reflect *both* the cognitive structure and the social structure of specialties. However, since this hypothesis will require much further elaboration and testing, it will only be stated here. A proper test would involve comparison of informal communication patterns with the bibliographic relations established by citations. Studies of this kind may also help unravel the relations between cognitive and social factors in the development of specialties and to determine the extent to which one is dependent on the other.

The importance of obtaining clusters that correspond to specialties can be illustrated by contrasting results here with alternative outcomes, which could have occurred but did not. First, it was possible that none of the highly cited papers in the *SCI* were co-cited. In that case, no clusters would have formed, and each highly cited paper would be an island unto itself. Second, all the highly cited papers could have been so strongly interrelated that only one gigantic cluster emerged, no matter how the clustering parameters were manipulated—the unity of science, forged presumably by interdisciplinary research, proving to be so powerful as to defy a breaking-up of knowledge into smaller subdivisions. But in fact, the outcome fell between these two extremes.

An important question is whether this outcome followed from the clustering algorithm applied. Practitioners of cluster analysis know that the outcome of a hierarchical clustering algorithm is as it was described in the previous paragraph—at one extreme a set of independent entities; at the other, a single cluster consisting of all the entities. The "natural" structure, if one exists, will depend on how the entities (in this case, cited documents) behave between the extremes—that is, whether they maintain a set of stable groupings over changes in level or strength of connection. Using a mountain range analogy, at the highest altitude the peaks are visibly separate; as the altitude is lowered, the peaks begin to merge into the mass; and finally, at the lowest level, all are united in a single mass. In terms of this analogy, science appears to consist of many sharply divided peaks which, for the most part, remain separate over a wide range in altitude. At very low levels, however, large merges occur, and eventually little, if any, of science remains unconnected at ground level. The point at which this merging begins is when co-citation reaches about 10% between papers (e.g., when one-tenth of all citations to the two documents are co-citations). At levels or altitudes only slightly higher than this (e.g., 15%), significant fragmentation remains.

The mountain range can also be more than an analogy, as shown by the following example. A cluster of six papers in particle physics (strong interactions) is depicted in Figure 1 as a contour diagram, with Feynman's 1969 paper on the parton model as the highest peak. This configuration or "landscape" was derived in a completely objective manner. First, each document was interpreted as a normal probability hill equal in volume to its citation frequency. Second, the distance between two hills was calculated by allowing the hills to interpenetrate until the volume of interpenetration was equal to the co-citation frequency. After all the distances were thus determined, multidimensional scaling was used to find that configuration of points in two dimensions that best fit the given distances. Finally, the contour lines were drawn so that along a given line the density of citing authors was constant.

The emergence of small groups of highly cited documents as the dominant pattern deserves special consideration, especially in terms of the clustering algorithm employed. The procedure is usually described as a single-link algorithm (23), because only one link of sufficient strength is required for membership in a cluster. This is, perhaps, the simplest of all clustering procedures, and it was dictated by the size of the data base being clustered—for example, typically about 10,000 highly cited documents. The single-link algorithm also has the weakest possible criterion for cluster membership, and it has been criticized for giving rise to "chaining"—the stringing together in a cluster of objects that bear little similarity to one another. Nevertheless, the application of single-link clustering to co-citation data does not exhibit this tendency, except at very low levels of co-citation: the average cluster size is only about five documents. Therefore, the small-cluster outcome appears to be intrinsic to the data rather than the result of the clustering methodology.

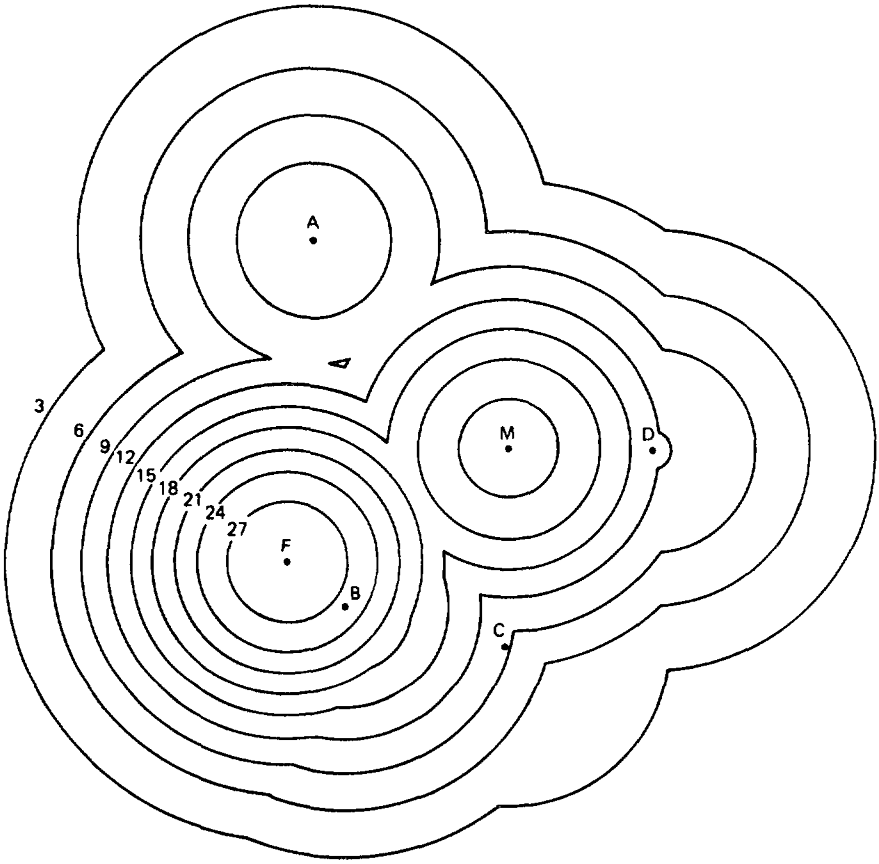


Figure 1. Contour map of six papers in particle physics. Each paper (designated by a letter code) is represented by a normal probability density function in the plane, with the volume of the hill set equal to the paper's citation frequency. Distances between the hills are determined by allowing the hills to overlap until the volume of overlap equals the co-citation frequency. The positions of the papers in the plane are determined by metric scaling (M-D-SCAL) on the set of 15 input distance. Along any contour line, the density of citing papers is constant and equal to the number on the line. *Key:* A = Amati, 1962; B = Benecke, 1969; C = Caneschi, 1969; D = DeTar, 1971; F = Feynman, 1969; M = Mueller, 1970.

The fact that clusters do eventually merge by chaining when the co-citation level is very low allows a further manipulation of the data, which provides a picture of how the specialties relate to one another. If, for example, a set of clusters is derived at a co-citation level of 11 (i.e., all documents co-cited 11 or more times appear in the same cluster), the clusters can be related by summing co-citation linkages between documents in different clusters of a strength of 10 or less—which would be, to use the mountain range analogy, low ridges connecting the sharp peaks. With the intercluster links (called "cluster co-citation"), "maps of science" can be constructed in terms of specialties. It is possible then to examine the way physics clusters relate to chemistry clusters and how the latter in turn relate to clusters in the biomedical sciences. Studies of this kind would not be possible if the *SCI* were not a multidisciplinary data base. This overall mosaic of specialties has important implications for studying the nature of interdisciplinary activity, since linkages between specialties of diverse subject matter indicate an exchange or a sharing of interests or methodology.

The map of biomedical clusters for 1972, shown in Figure 2, was derived by the application of two thresholds: No cluster with fewer than five documents is included, and must be linked with another by a co-citation strength of 100 or more. A linkage between two clusters means that a number of authors are citing documents in both clusters and thereby creating intercluster co-citations. Such citations should reflect the degree of interdependence of one specialty on another or the extent of interdisciplinary effort.

The map shows four major regions corresponding to major research areas in present-day biomedical research: Chromosomes and RNA virus work (viral genetics), at the upper left; immunology, upper right; biological membranes, lower right; and cyclic AMP, lower left. A 1972 specialty now known to have been a key growth point is "microtubule protein" (in the lower left-hand corner), which developed linkages with all major areas on the map in 1973.

The goal implied by the term "mapping science" is most nearly realized by the use of cluster co-citation in dealing with links between specialties rather than links between documents. Many of the document-level techniques (e.g., graphing, multidimensional scaling) can also be applied to the specialty level. The final step in achieving the goal of mapping science is to relate large disciplinary units (biomedicine, physics, chemistry) to one another. The specialty clusters exhibit some tendency toward hierarchical or nested structure—that is, for clusters in chemistry to merge at lower co-citation levels and to form a macrocluster. This tendency may be taken as evidence for a larger structure by discipline, in which the specialties are ordered.

The present system takes advantage of that assumption to arrive at a large-scale map of science. The documents are first clustered at a very low level (e.g., 7), which has the effect of grouping some, but not all, of the specialty clusters into disciplinary clusters. Then cluster co-citation is determined for the disciplinary clusters. A "map of science" that includes some of the largest clusters at Level 7 is shown in Figure 3. Nonmetric multidimensional scaling of cluster co-citation frequencies was used to determine the configuration.

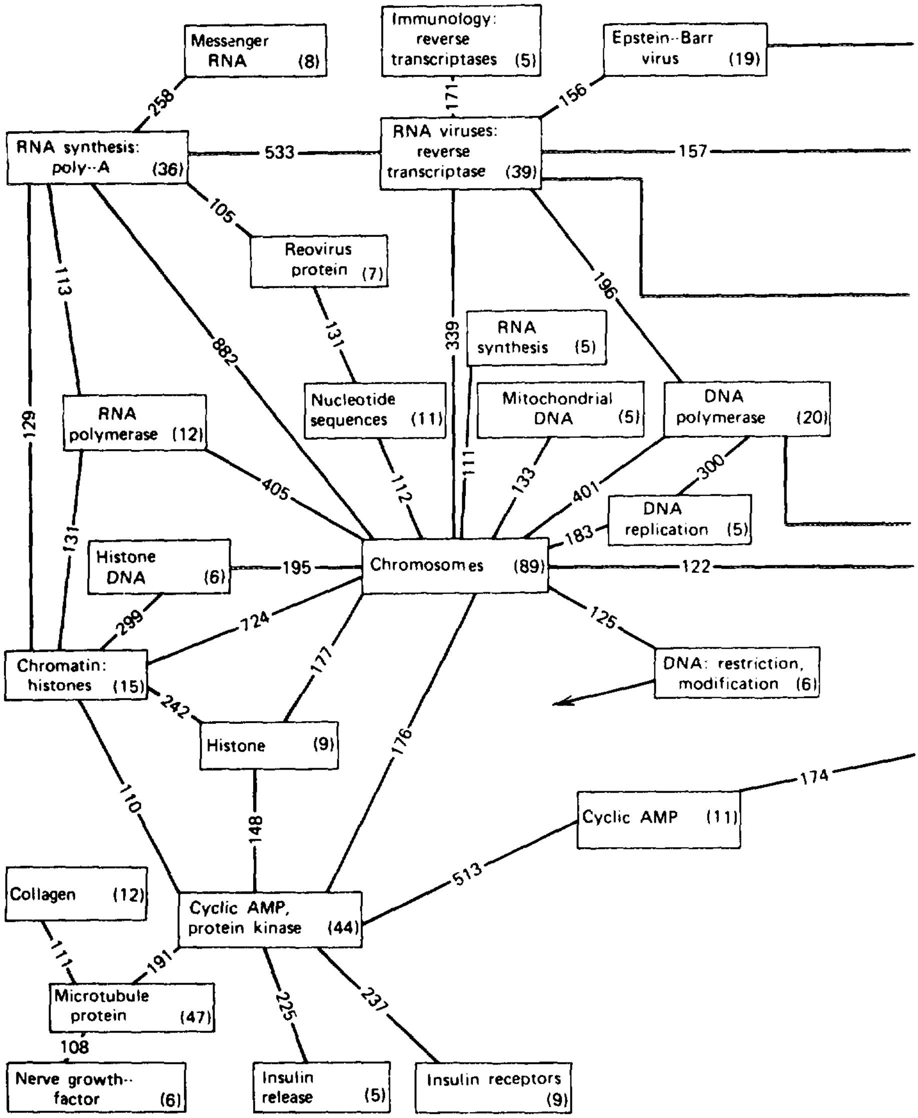
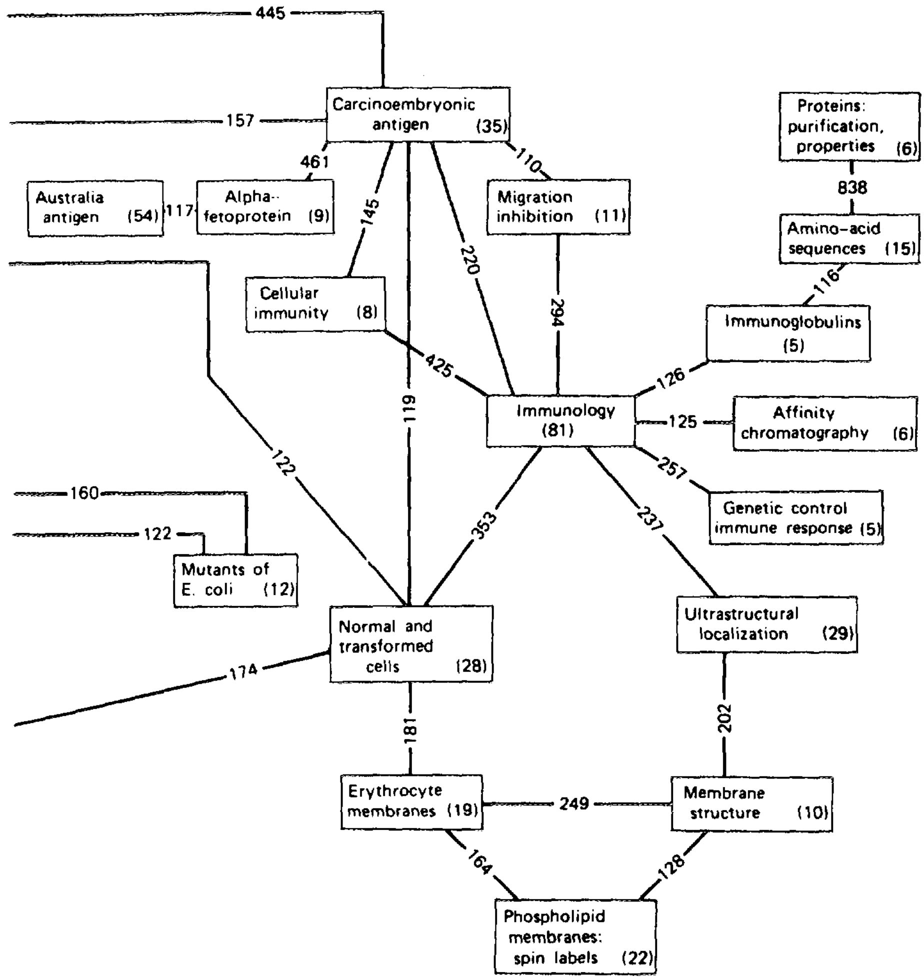


Figure 2. This is a map of major biomedical clusters in 1972. Each box is a cluster of highly cited documents (name of cluster and number of documents given in each box) obtained from the 1972 *Science Citation Index* at thresholds of 10 citations per document and 11 co-citations per docu-



ment pair. In constructing this diagram two further thresholds were used: Clusters were connected by lines only if there were at least 100 co-citations between documents in the two clusters (see number on each line), and only if the clusters contained at least five highly cited documents. Reprinted with permission of *Mosaic* magazine. Originally appeared in S. Aaronson, "The Footnotes of Science," *Mosaic*, **6**, 22–27 (1975).

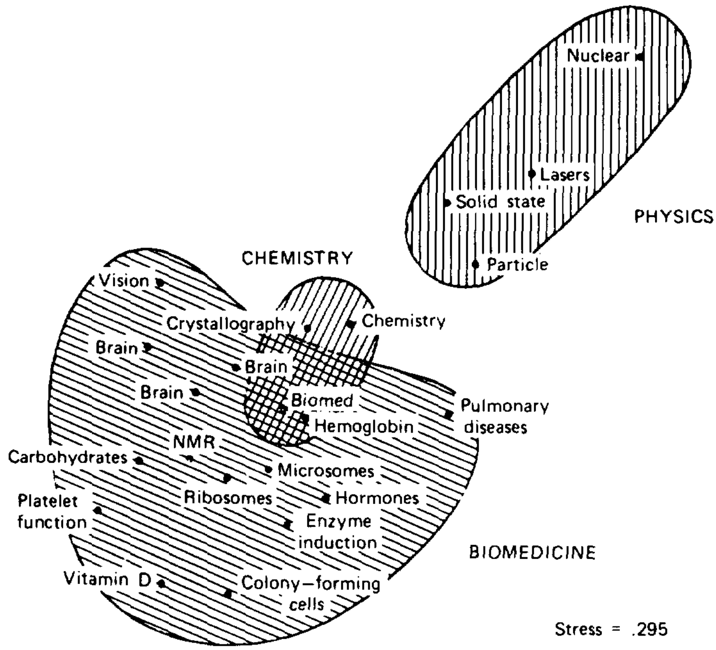


Figure 3. Major disciplinary clusters obtained at Level 7 from the 1973 *Science Citation Index* and containing 30 or more highly cited papers. A matrix of cluster co-citations is constructed and the matrix is used as input to Kruskal's multidimensional scaling (M-D-SCAL), using its nonmetric option. Stress is a measure of goodness-of-fit of the input similarities to the final configuration. The configuration of specialities shows three regions: a physics region, a biomedical region, and a chemistry region that is between physics and biomedicine and partially embedded in biomedicine.

The notion of mapping implies dealing with objects or entities that have a location in a space of some number of dimensions in which the distance between objects is meaningful and well-defined. Indeed, the language used in talking about science is filled with spatial metaphors. Physics, chemistry, and biology are "fields," related logically, socially, or in terms of shared subject matters; one field is said to have a "close bearing" on another. Information scientists have long posed the problem of finding a measure of distance between classification headings, and a library or a classification scheme is set up (theoretically) to make related subjects physically close to one another. Mapping science is an attempt to arrive at a physical representation of fields and disciplines—and, at a lower level, of individual papers and scientists—in which the relative locations of entities is depicted.

But do the maps of science derived from citation data have reality in this strict spatial sense? Measures of association derived from citation data do not necessarily imply the existence of a metric space. Spatial representations can be obtained merely from ordinal data, as shown by multidimensional scaling. The ontological status of maps of science or other cognitive maps will perhaps remain speculative until more has been learned about the structure of the brain itself. Whatever their physical reality, maps of science are certainly useful as heuristic tools. The same might be said of the "mental maps" being constructed by geographers (24).

MODELS OF SCIENTIFIC CHANGE

Up to this point, clusters and maps of clusters have been examined in terms of a single source year. But it is also possible, since *SCI* is a multiyear file, to examine shifts in clusters over time and to investigate the nature of change in specialties. The problem of change is extremely complex, and we can hope only to point to certain directions in which research must proceed. Historians and philosophers of science have long regarded development and change in science as a central problem, and a number of theoretical statements have emerged that have far-ranging implications, most notably those of Kuhn (13) and Toulmin (25). In fact, Toulmin's recent book contains a set of diagrams that resemble quite closely "historical maps" derived using citation relationships.

Moreover, Toulmin's distinction between intellectual "disciplines" and "professions" parallels our distinction between clusters of cited documents and corresponding clusters of citing authors. The model Toulmin proposes is an adaptation of evolutionary theory and may be directly testable by using citation data. The test might involve identifying his "variants" of an idea among the papers citing a particular seminal work and then seeing which, if any, of these works becomes highly cited itself. In this interpretation, high-citation frequency is evidence for what Toulmin calls "selection" of an idea, and variation occurs when authors cite a previously "selected" idea. Toulmin employs a sort of replacement or gradual-improvement model of science, while Kuhn employs a model in which sudden shifts of perspective alternate with periods of relative stability. Kuhn's model can also be formulated in citation terms and tested.

The manner in which clusters of cited documents change over time should be examined. Periods of stability or gradual turnover should occur when the document clusters continue and maintain a constant configuration. There should also be sudden changes (perhaps in the course of a single year) in the configuration and the document set, with very little overlap between succeeding years. Even a third model is conceivable (perhaps to be called the Popper model [26])* in which a continuing "revolution" in all specialties is evidenced by a large and continuous turnover of highly cited documents in the cluster. We do not attempt a systematic test of the models here; but we do offer some evidence to suggest that the empirical situation is highly complex, and examples can be found to suit all the above theories.

The technique for studying change with citation data and clustering methodology is to identify corresponding clusters in successive annual files. This can be done easily, because at least some of the highly cited papers in each cluster persist from year to year. At the same time, the entrance of new highly cited documents and the exit of old ones can be seen. The core of continuing documents also changes its configuration because of weakening or strengthening of selective linkages. These changes may be depicted and analyzed by the topological approach (graphic theoretic methods) or the spatial approach (multidimensional scaling techniques). The topological approach focuses on the cluster as a formal graph (nodes and edges) where all linkages above a minimum threshold are considered equal. The spatial approach employs an ordination technique (e.g., multidimensional scaling) to assign positions in *N*-dimensional space to each of the documents (27). The study of change then involves tracking the documents' motion through the space over time. Of the two methods, the second (spatial) approach appears more powerful: It makes use of more of the information and it appeals to a body of highly sophisticated statistical techniques. Graphing, however, can be useful in gaining a qualitative impression of how the structure is changing. These patterns of change are complicated by mergers, sometimes large-scale, or splits in clusters from one year to the next. Such changes probably reflect boundary shifts between specialties—for example, a new subspecialty breaking off from its parent specialty or two previously separate specialties joining forces, perhaps to attack a common problem.

The sociometric and historical literature on specialty development (28) emphasizes the rapidity with which changes can occur in the early phase of development, when growth can be exponential. Recent research suggests that some biomedical specialties can emerge in as little as six months after the publication of the discovery papers (29). Our clustering studies generally confirm this potential for extremely rapid growth, although not all specialties follow the pattern. Studies of specialties also reveal leveling-off periods and periods of decline. These results suggest that specialties go through various

*Popper apparently believes that science is, or at least should be, continuously revolutionary: "In my view the 'normal' scientist, as Kuhn describes him, is a person one ought to be sorry for" (p. 52).

phases of development from birth to death—a kind of life cycle. Thus far, cluster data have not revealed such a deterministic pattern but the following composite picture has emerged from a number of cases.

From one to three highly cited and highly co-cited papers appear as a tightly knit cluster in the first year. These papers contain simultaneous or successive discoveries quickly recognized as breakthroughs. Occasionally, an old paper included in this initial "discovery" group may have the formal role of precursor. Usually, at least one of the papers is very recent (perhaps only one year removed from the source year). Often the discovery group is weakly tied to an older cluster of methodological importance. One might hypothesize that the old cluster provides a legitimizing context for the new ideas or perhaps a source of manpower for the new specialty. After one year the nucleus of discovery papers expands dramatically, and sometimes explosively, to include perhaps severalfold the number of highly *cited* papers—not discovery papers but early working-out papers that have exploited the new ideas. Since they are closely dependent on the discovery papers, they are all highly co-cited with the discovery papers. Concurrent with the rapid growth in the new cluster, the old methodological cluster disappears or declines in importance.

The third or fourth years display increasing stability, although that stability can be short-lived. This middle period may involve, for example, the appearance of review papers in the cluster or movements into applied science or technology. Some of the original working-out papers disappear and are replaced by more recent up-to-date papers. The distribution of cited papers by publication date settles down to the average for all of science, with the mode at about two or three years before the source year. In most cases the original discovery papers persist as highly cited papers in the cluster and provide a kind of framework for later developments. Stability is often followed by decline, which can be manifested in a disintegration of the cluster into smaller fragments. Any dramatic novelty in the specialty would initiate a new sequence of events, similar to the one just described, in which the role of the methodological cluster would be played by the original cluster. Apparently, specialties must face either eventual demise or substantial transformation to incorporate new material.

Much of our discussion must be regarded as hypothesis based on examination of numerous examples; we do not wish to convey the impression that specialty development can be fitted into a neat predictive theory. A four-year (1970–1973) study of some of the 31 continuing specialty clusters shown in Figure 2 yields the percent-change results listed in Table 1. The overall mean rate of document continuation is about 55% for these clusters, but the variation in the percentage can be large. The change from one year to the next in the same specialty varies from gradual to dramatic. In the

Table 1. Percent Change in Sample of 31 Continuing Clusters, 1970–1973
c = continuing, *d* = dropping, *n* = new documents

Specialty	Direction of Change	1970–1971 (%)	1971–1972 (%)	1972–1973 (%)
Nuclear levels	c	58	45	25
	d	21	55	17
	n	21	0	58
Adenosine triphosphatase	c	67	25	67
	d	0	50	22
	n	33	25	11
Australia antigen	c	55	54	57
	d	4	26	30
	n	41	20	13
Proton-proton elastic scattering	c	50	7	44
	d	50	21	25
	n	0	72	31
Ultrastructure of secretory cells	c	50	43	60
	d	12	57	0
	n	38	0	40
Nuclear magnetic resonance	c	37	55	23
	d	13	9	54
	n	50	36	23
Polysaccharides	c	46	44	36
	d	46	34	7
	n	8	22	57
Crystallization of polymers	c	100	100	100
	d	0	0	0
	n	0	0	0
Affinity chromatography	c	60	67	72
	d	20	0	14
	n	20	33	14
Leukocytes: chronic granulomatous disease	c	40	63	33
	d	13	5	53
	n	47	32	14
Collagen	c	80	40	27
	d	20	0	40
	n	0	60	33
Erythrocyte membranes	c	9	15	58
	d	64	5	42
	n	27	80	0
Delayed hypersensitivity	c	77	46	50
	d	15	27	29
	n	8	27	21

Table 1. *(continued)*

Specialty	Direction of Change	1970–1971 (%)	1971–1972 (%)	1972–1973 (%)
Fission of deformed nuclei	c	45	63	36
	d	22	25	7
	n	33	12	57
Malignant hyperpyrexia and hyperthermia	c	29	45	70
	d	14	22	0
	n	57	33	30
Transfer RNA	c	52	29	67
	d	0	71	33
	n	48	0	0
Crystallography	c	33	43	40
	d	11	26	20
	n	56	31	40
Subacute sclerosing panencephalitis	c	60	80	66
	d	20	0	17
	n	20	20	17
Marek's disease	c	33	42	12
	d	27	50	63
	n	40	8	25
Tumor-specific immunity	c	100	100	100
	d	0	0	0
	n	0	0	0
Solid state: disordered systems	c	75	44	64
	d	0	0	18
	n	25	56	18
Hepatic porphyria	c	43	22	33
	d	57	11	56
	n	0	67	11
Immunoglobulin-A	c	50	38	75
	d	39	46	12
	n	11	16	13
Spectrophotometric studies of complexes	c	80	100	100
	d	20	0	0
	n	0	0	0
Myocardial contractility	c	31	20	25
	d	38	33	58
	n	31	47	17
Virus-specific proteins	c	50	80	100
	d	17	20	0
	n	33	0	0

Table 1. *(continued)*

Specialty	Direction of Change	1970–1971 (%)	1971–1972 (%)	1972–1973 (%)
Plasma hormones	c	29	50	27
	d	28	33	9
	n	43	17	64
Magnetic properties of alloys	c	100	100	100
	d	0	0	0
	n	0	0	0
Lesch-Nyhan syndrome	c	60	21	26
	d	40	0	67
	n	0	79	7
Multidimensional scaling	c	100	100	100
	d	0	0	0
	n	0	0	0
Pseudopotentials	c	40	50	34
	d	53	37	8
	n	7	13	58
Mean percent	c	56	53	56
	d	21	21	22
	n	23	26	22

four-year period, about one-third of the 31 specialties experienced major shifts in the set of cited documents—that is, all but one or two of the cited documents in the cluster dropped out, and an almost entirely new set appeared. If dramatic shifts of this kind can be correlated with the occurrence of revolution in specialties, we might hypothesize that a specialty will undergo, on the average, one revolution every 12 years. In this regard, the percentage of continuing documents in a cluster is a good indicator of whether a revolution is occurring.

A common feature in the 31 cases studied was the way in which change occurred: Documents moved in and out of clusters in groups rather than singly, and entire clumps of documents would disappear in one year and be replaced by new clumps the next year. We can hypothesize that these changes represent shifts in the leadership of the specialty from one school or group to another.

An example of a specialty in this sample of 31 cases that has undergone a major shift over the period 1970–1973 is research on the protein collagen, the major component of all connective tissue. Comparing the 1970 and 1973 networks (see Figure 4) reveals that there are no papers in common: The specialty has undergone a complete shift in the cited-document set. The shift is clearly evident in the 1972 diagram, with the appearance of a subcluster of five 1971 papers centering around the works by Bellamy and Layman. Here we are dealing with a multiple discovery, which occurred in 1971 and became evident in citations the following year, of a new substance called "procollagen," the biosynthetic precursor of ordinary collagen. The papers by Layman and Bellamy are cited by other 1971 papers as those first announcing the finding. The new subcluster is also attached to the old collagen cluster through Piez's 1963 paper, which is primarily of methodological importance. In 1973 vestiges of the old collagen cluster disappear, and the cluster consolidates around the new 1971 work.

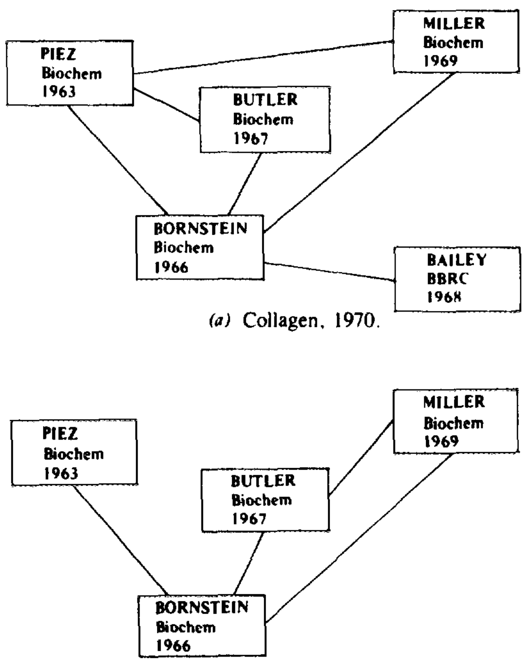
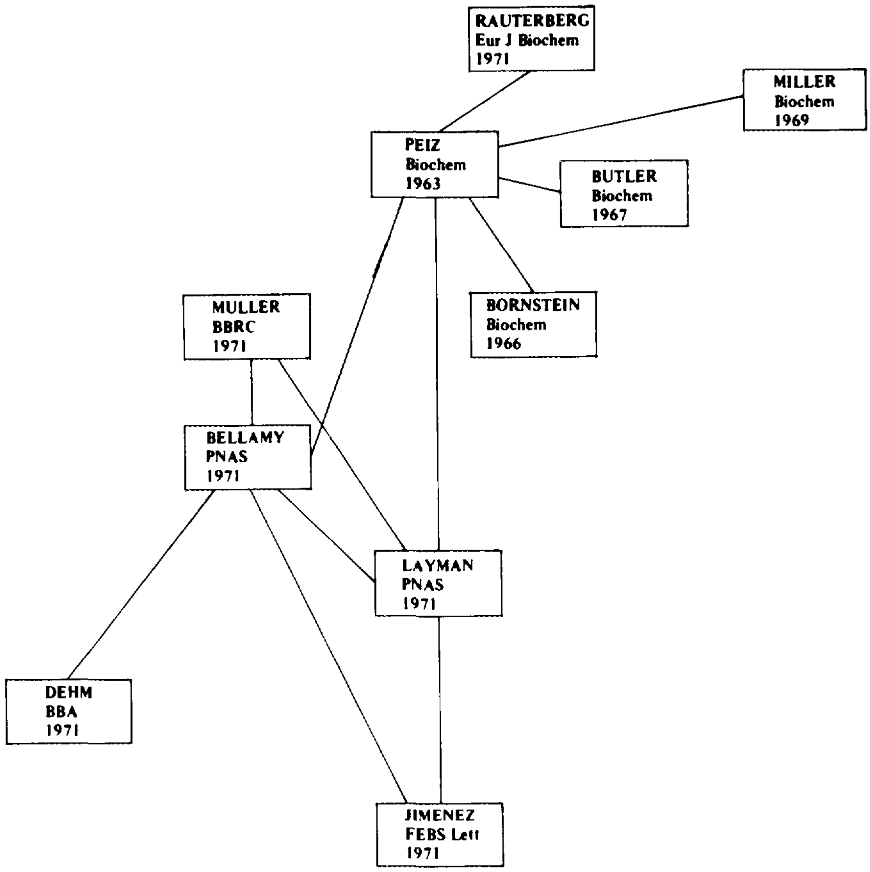
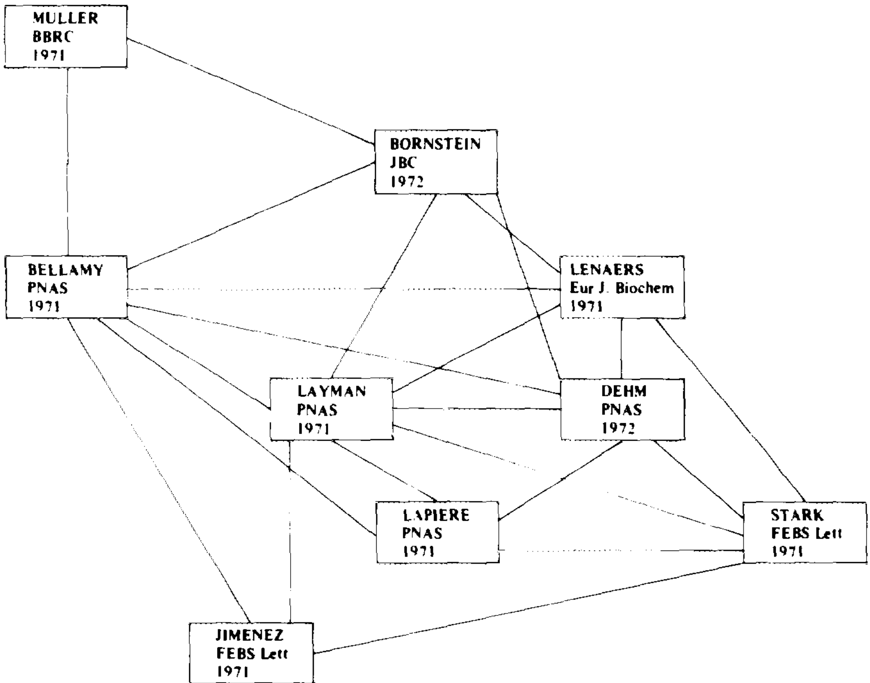


Figure 4. Development of a specialty cluster, 1970–1973 (citations to journal articles). The figure shows the evolution of the collagen cluster over the four-year period 1970–1973. Boxes contain the names of first authors of the highly cited papers and years of publication. Lines connect papers co-cited at least 11 times in the corresponding source year. *(a)* Collagen, 1970. *(b)* Collagen, 1971. *(c)* Collagen, 1972. *(d)* Collagen, 1973.



(c) Collagen, 1972.

To validate this picture of the development of the collagen specialty, the citing authors, who were assumed to be specialists currently working in this area, were asked to respond to a questionnaire. The specialists surveyed were not shown any of our results but were only asked to answer a series of questions such as: What are the most important scientific advances or developments in collagen research in the past five years? and What papers were the first to describe these advances? All respondents so far have named the discovery of procollagen as one of the most important advances in the past five years. Of the nine papers said to contain the important discoveries, five are in the cluster, one is in a neighboring cluster (on genetically different collagens), one is not in any cluster, and two are 1973 papers that could not



(d) Collagen, 1973.

be expected to appear in the cluster until 1974. The specialists were also asked whether collagen research had undergone a conceptual shift in the past five years. So far, all of the specialists have responded affirmatively, giving similar reasons for the shift. Although it would be difficult to characterize the change in this specialty as a "revolution" in the Kuhnian sense, a major redirection of research has clearly taken place and the specialty has been revitalized by the infusion of new ideas.

These discussions of revolutions in science and life cycles for specialties may seem to be contradictory: How can a specialty be subject both to periodic upheavals and to gradual and orderly progression from birth to death? The reconciliation of these views is consistent with the data at hand and constitutes an hypothesis, deserving further attention, on the development of specialities. In brief, we suggest that revolutions are most likely to occur at the end-point of specialty development—for example, when a state of decline or stability has been reached. What emerges after the major shift in concepts is an essentially new specialty having certain vestigial links with the past, as symbolized by older papers that persist through the revolution.

From Basic to Applied Research

The use of co-citation methodology for the specialty of amorphous semiconductors (30) exemplifies the identification of research trends from basic to applied. After deriving the cluster of cited documents, the citing papers on the subject were used to identify persons and organizations responsible for the research and to detect a movement toward practical application of those devices which had been mainly of basic-research interest.

The co-citation patterns for the period 1968–1972 were investigated. It was observed in 1973 that a new group of papers tied to the previous group began to appear, and that the new group was related to applications of earlier research. This was determined through examination of the titles in the citing and cited articles. Further examination showed that, although all but one of the previous group worked at universities, all the new authors worked in industrial organizations (see Kochen this volume). Evidently, there was movement toward application in the 1972–1973 period, as compared to basic research in the 1968–1971 period.

The examples given, of course, only illustrate some of the capabilities of the methodology and indications for future research; they do not prove or disprove any of the theories of change discussed. What exactly would constitute a confirming or refuting instance is not clear. But it is clear that changes in scientific theories must first be reinterpreted in citation terms—that is, how *do citation patterns change in response to social or conceptual* change in specialties? This may involve redefining or narrowing some conceptions. It also involves finding appropriate "correspondence rules" between the theoretical constructs and observable citation data. For example, does the appearance of a new cluster of highly cited papers correspond to the cognitive and social event of the emergence of a new specialty? One method of validation is to survey the opinion of the specialists, as was done in the case of collagen. Another would be to examine what science writers regard as newsworthy developments in science and compare their selections with specialty clusters. This method has been pursued in an informal way for news items appearing in *Physics Today* (see Holton, this volume) with resulting agreement between their selection and what appear to be our "hottest" physics clusters. Another method would be to determine the correspondence between informal patterns of communication and citation networks. The development of such correspondence rules is essential if citation or other bibliographic data are to be used in sociological or historical research.

However, it would be wrong to require agreement of citation data with more traditional techniques of investigation. The problem of validating citation data is more complex than that: It involves a new tool or instrument (a citation index), which provides a new perspective to be interpreted. We must learn to do that and then to relate the new information to what is already known. In other words, the interpretations of citation data, not the body of data, are what need validation.

Another frequent question is "Can we use citation data in science forecasting?" For example, can any property of the current specialty structures discussed be used to fortell its future configuration? We do not yet know whether research on the life cycles of specialties will turn up any consistent patterns of change that would be of predictive value. The powerful potential for change contained in the unanticipated discovery is quite apparent. But sudden and dramatic shifts occur in clusters from one year to the next (e.g., in the collagen specialty); and to predict where these will occur on the scale of relationships between specialties—where single events may have less impact—is beyond current capability. We could look for trends in interdisciplinary linkages, such as immunology moving to link with cyclic AMP or biological membrane work moving to link with viral genetics. On the largest scale we might look for corrections forming between disciplines—for example, physics forging links with biomedicine.

Clusters as Science Indicators

We should also ask in what sense are the clusters to be considered a form of science indicators? A cluster of highly cited papers is perhaps an indicator of consensus (31)—at least in the sense that a number of researchers, by focusing their attention on a narrowly defined problem, implicitly agree that it is a worth-while object for their attentions.

The cluster is an indicator of this focusing of attention by the community (see Cole et al., this volume). It points to the problems scientists regard as important and of immediate priority. Hence, clusters are specialty indicators: They provide the information that a certain number of scientists are directing their attention to research on the Epstein-Barr virus or plate tectonics, and they indicate whether that activity is related to work in any other specialties. Furthermore, something can be discerned about the current rate of change in the specialty—whether it is undergoing a revolution or moving through a period of stability. Presumably, this is the kind of information needed to ascertain how specialties are progressing—that is, whether some need revitalization or could use the stimulus of support—according to external measures of priority having their origin in the larger society (see Ezrahi, this volume).

At present, a fully computerized system exists at ISI for clustering annual cumulations of the *SCI* in an off-line mode. Another project currently under way is the development of a fully on-line system (now in the horizon stage) for performing many of these functions. In its final form, it will allow on-line access to the data base in both linear (teletypewriter) and two-dimensional (graphical display) modes (see Kochen, this volume).

The system will automatically structure large fields into clusters, so that at each hierarchical level a manageable amount of information is presented. The researcher could then select substructures to analyze, and the system would automatically proceed to the next level and once again display a manageable amount of information. At the lowest level the system would display the direct citation relations between the various elements of the field or specialty being studied. For example, the researcher might start with a general query on high-energy physics. The graphic system would display the half-dozen or so major clusters in that field, and the user could then select one of the subclusters on particle interactions. The next level might consist of subclusters on strong interactions, weak interactions, and other areas. Finally, by the selection of one of the subclusters, the system will respond with a network of papers in the user's selected area.

Perhaps not so basic in research orientation as the mapping of science but still important in relation to developing indicators are studies—those already carried out and those planned for the future—that use citation data for determining the achievement or impact on science of individuals and organizations. In these studies data from the *SCI* are used to provide quantitative measures of impact. The citation data obtained for the study sample are then used to establish the relative standing of individuals or institutions. The studies are not intended to rate or grade individuals or institutions without regard to other factors, since the major purpose is to provide an indication of impact and not an absolute measure. As more and more of these studies are undertaken, extensive data are being collected that will be used in developing techniques for establishing confidence limits for the results. Methods for integrating the variables associated with the citation data obtained are also being studied.

The questions being addressed in these efforts are: Is a citation count alone a sufficient indicator of impact? Is the average number of citations per paper or per department a better indicator than simple citation counting? What is the relation between an individual's age or the date of his Ph.D. and citation patterns? What normalizations must be established for different age-groups?

Studies of the characteristics of papers cited with high, average, and low frequency will help determine more exactly the relation between the content of a paper and citation data (see Kochen and Cole et al., both papers this volume). Finally, studies of citation patterns of different fields of science should help to normalize for this variable. Certain facts concerning citation characteristics of different fields are already known—for example, the chemical-physics articles, on the average, cite about 20 papers compared to fewer than 10 for mathematics papers. However, there are more chemical-physics papers than mathematics papers to cite—a point often forgotten when people worry about the problems as Janke did (32).

This paper has centered on the application of bibliometrics—in particular, citation metrics. The literature of science, as a by-product of scientific work and sometimes as the culmination of that work, has great potential for the study of science and for developing indicators of the condition of science.

The research directions described should provide indicators for measuring the degree of scientific activity, the quality of research, and scientific achievements. These measures may also have value in identifying the scientist's and the public's options in regard to support for mission-oriented research as opposed to basic research. With such means available, we should be able to deepen our understanding of the problem of setting priorities in relation to societal goals (see Ziman, this volume).

This issue of balanced support of research is also important. Indicators that can be derived from citation analysis could help to identify underdeveloped or currently neglected areas of science. Measuring the degree of research activity in such areas as mental illness and drug addicton, for example, should generate good information to help us decide whether, in terms of our sense of priorities, we have allocated resources correctly.

Citation data can also be used as a measure of national and international science activity. The computer file can be expanded to include the addresses of both citing and cited authors, thus permitting measurements of dependence and independence by individuals and countries. Such analysis would be much more useful then the simple tabulation of the contributions of different countries to various fields presented in *SI-72*. What needs to be measured is international exchange of scientific ideas.

Finally, as the term *indicator* implies, we must be concerned with the evolution of systems over time and the sampling and measuring of systems at successive points in time. Citation data probably will afford little material for the advanced futurologist, but short-term extrapolation may be feasible. The life cycles of specialties must be studied—from their emergence as small clusters of highly cited discovery papers through the explosive initial phase of growth to the stabilization of patterns and the eventual decline or revitalization. Such studies have ramifications beyond an increase in general knowledge of how specialties are born and mature. Perhaps the indicators derived from such study will help us to foresee the need for new journals and books (33). Studies of the new terminology associated with the specialty will aid in controlling for the literature, thus permitting better anticipation of the words and phases to use in retrieving the new literature.

The ability to study change in science—provided by citation data—is an important contribution that could be even greater if an *SCI* were available for the period 1900–1960. With such a compilation, we would have a continuum for the entire twentieth century, and the sociological and historical studies that are so important to the basic question with which we are struggling could be greatly advanced.

REFERENCES

1. E. Garfield, "Citation Indexing for Studying Science," *Nature*, **227**, 669–671 (1970).
2. W. O. Hagstrom, "Inputs, Outputs, and the Prestige of American University Science Departments," Sociological Education, **44**, 375–397 (1971).
3. A. Pritchard, "Statistical Bibliography or Bibliometrics?" *Journal of Documentation*, **25**, 358–359 (1969).
4. J. R. Cole and S. Cole, *Social Stratification in Science*, The University of Chicago Press, Chicago, 1973.
5. D. J. de Solla Price, "Networks of Scientific Papers," *Science*, **149**, 510–515 (1965).
6. E. Garfield, I. H. Sher, and R. J. Torpie, *The Use of Citation Data for Writing the History of Science*, Institute for Scientific Information, Philadelphia, 1964, 86 pp.
7. *Science Citation Index 1973: Guide and Journal Lists*, Institute for Scientific Information, Philadelphia.
8. E. Garfield, "Citation Analysis as a Tool in Journal Evaluation," *Science*, **178**, 471–479 (1972).
9. H. Small, "Characteristics of Frequency Cited Papers in Chemistry," Final Report NSF Contract C-795, September 1974.
10. A. Einstein, "*Zur Elektrodynamik Bewegter Korper*," *Annalen der Physik* **17**, 891–921 (1905).
11. M. Polanyi, *Personal Knowledge*, Harper & Row, New York, 1958.
12. "Exit polywater," *Scientific American* **229**, 66 (1973).
13. T. S. Kuhn, *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 1962, p. xi.
14. O. H. Lowry, N. J. Rosenbrough, A. L. Farr, and R. J. Randall, "Protein Measurement with the Folin Phenol Reagent," *J. Biol. Chem.* **193,** 265 (1951).
15. N. Bohr, "On the Constitution of Atoms and Molecules, Part I, *Philos. Mag.* **26**, 1–25 (1913).
16. R. K. Merton, "Singletons and Multiples in Scientific Discovery," *Proc. Amer. Philos. Soc.* **105** (5), 470–486 (1961). Reprinted in *The Sociology of Science: Theoretical and Empirical Investigations*, The University of Chicago Press, Chicago, 1973, pp. 343–370.
17. K. Sparck-Jones and D. M. Jackson, "The Use of Automatically Obtained Keywork Classifications for Information Retrieval," *Information Storage and Retrieval* **5**, 175–201 (1970).
18. H. Small, "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents," *Journal of the American Society of Information Science* **24**, 265–269 (1973); I. V. Marshakova, "System of Document Connections Based on References," *Nauchno-Teknicheskaia Informatsiia*, **2**, 3–8 (1973).

19. M. M. Kessler, "Bibliographic Coupling between Scientific Papers," *American Documentation*, **14**, 10–25 (1963).

20. H. Small and B. C. Griffith, "The Structure of Scientific Literatures I: Identifying and Graphing Specialties," *Science Studies*, **4**, 17–40 (1974).

21. B. C. Griffith and N. C. Mullins, "Coherent Social Groups in Scientific Change," *Science*, **177**, 959–964 (1972).

22. D. Crane, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, The University of Chicago Press, Chicago, 1972.

23. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, Freeman, San Franciso, 1973, pp. 188–308.

24. P. Gould and R. White, *Mental Maps*, Penguin Books, Harmondsworth, England, 1974. See also Garfield, Sher and Torpie, *op. cit.*

25. S. Toulmin, *Human Understanding: The Collective Use and Evolution of Concepts*, Princeton University Press, Princeton, 1972, pp. 201, 203, 205.

26. K. Popper, "Normal Science and Its Dangers," in *Critism and the Growth of Knowledge*, I. Lakatos and A. Musgrave, eds., Cambridge University Press, Cambridge, 1970, pp. 51–58.

27. B. C. Griffith, H. Small, J. A. Stonehill, and S. Dey, "The Structure of Scientific Literatures II. Toward a Macro- and Microstructure for Science," *Science Studies*, **4**, 339–365 (1974).

28. W. Goffman, "A Mathematical Method for Analyzing the Growth of a Scientific Discipline," *Journal of the Association for Computing Machinery*, **18**, 173–185 (1971); G. Magyar, "Bibliometric Analysis of a New Research Subfield," *Journal of Documentation*, **30**, 32–50 (1974).

29. B. C. Griffith, "On the Nature of Social Science and its Literature," unpublished paper.

30. A. E. Cawkell, "Search Strategy, Construction and Use of Citation Networks, with a Socio-Scientific Example: Amorphous Semi-conductors," and S. R. Ovshinsky, *Journal of the American Society for Information Science*, **25**, 123–130 (1974); "Mapping Science by Co-citation Networks," unpublished report.

31. J. M. Ziman, *Public Knowledge: The Social Dimension of Science*, Cambridge University Press, Cambridge, 1968.

32. N. C. Janke, *Science*, **182**, 4118, December 21, 1973; and reply by E. Garfield.

33. E. Garfield, "Citation Analysis of Pathology Journals Reveals Need for a Journal of Applied Virology," *Current Contents*, No. 3, 5–8, January 17, 1973.

Offprints From
Toward A Metric Of Science:
The Advent of Science Indicators
Edited by Yehuda Elkana, Joshua Lederberg,
Robert K. Merten, Arnold Thackray, and
Harriet Zuckerman
Published by John Wiley & Sons, Inc.
Copyright Date 1978