Annals of Biomedical Engineering, Vol. 14, pp. 97-108, 1986
0090-6964/86 $3.00 + .00

MAPPING THE WORLD OF BIOMEDICAL ENGINEERING: ALZA LECTURE (1985)

Eugene Garfield

Institute for Scientific Information,® Inc.
3501 Market Street
Philadelphia, PA 19104

The key research areas of biomedical engineering are identified and analyzed. It is demonstrated how biomedical engineering fits into the world map of science. An analysis of biomedical engineering core journals provides statistical data about citation patterns in this discipline.

Keywords *— Cluster mapping, Scientometrics, Biomedical engineering, Citation index.*

The Institute for Scientific Information® (ISI®) has developed a system over the last 15 years that uses the data recorded in the Science Citation Index® (*SCI*®) to identify and map the various areas of science. We can do this exercise for almost any field that we choose. For the Biomedical Engineering Society's 1985 Alza Lecture, we identified the biomedical engineering literature related to the interests of that organization's audience. I think you will see that we can do the same basic exercise for any subject.

To give you a feeling for the size of the database that we're working with, the number of published items indexed in the *SCI* over the past 30 years grew from 80,000 in 1955 to nearly 565,000 in 1984. These half-million 1984 publications contained over 9 million cited references. The whole *SCI* file now contains nearly 150 million such cited references.

The main purpose of the *SCI* is to tell the researcher where a particular paper has been cited. However, the *SCI* actually consists of four distinct kinds of index: the *Citation Index*, the *Source Index*, the *Permuterm® Subject Index*, and the *Corporate Index*. Each of these gives you a different access point to the scientific literature. The terms "Citation Index" and "Source Index" are idiosyncratic to ISI's various citation indexes. The *Source Index* is an author index similar to the author indexes produced by *Chemical Abstracts, Biological Abstracts*, and so on. But the *Citation Index* is also an "author index," and to make the distinction, we've given them separate names.

To illustrate how the *Citation Index* works, consider the work of "Alberto Zaffaroni." While the *Source Index* tells you what he has published in a particular year, the *Citation Index* tells you which of his works have been cited that year. No matter how old the original article may be, it is essential that any new researcher using that article determine who has cited it and where, so as to know its current validity or use.

In the *Source Index* you will find a complete bibliographic entry for each paper by every author for the period covered. A major use for the *Source Index* is to find the complete title for a citing paper noted in abbreviated form in the *Citation Index* section.

Another approach to indexing this material is found in our *Permuterm Subject Index*. There are many traditional subject indexes that use subject headings from a standardized thesaurus. The *SCI*, however, uses natural language—the actual language used by authors. We permute the words in the title of each paper we index to create all of the possible pairs for that title. If researchers don't know an exact reference to begin with, they can use this subject index to locate a starting reference.

The fourth, or *Corporate*, index tells the researcher what has been published from a particular institution. If you look in the *Corporate Index* under "California, Palo Alto," you will find, for example, the Alza Corporation (2). There you will find the articles published by authors from this organization. The *Corporate Index* enables you to determine what a particular research organization is emphasizing.

Of the 9 million references that we processed in 1984 in the *SCI* there were 4,400,000 unique items that were cited an average of two times that year. (The ratio of citations in any one year to the number of published items cited in that year is frequently called Garfield's constant, although it isn't really constant (1).)

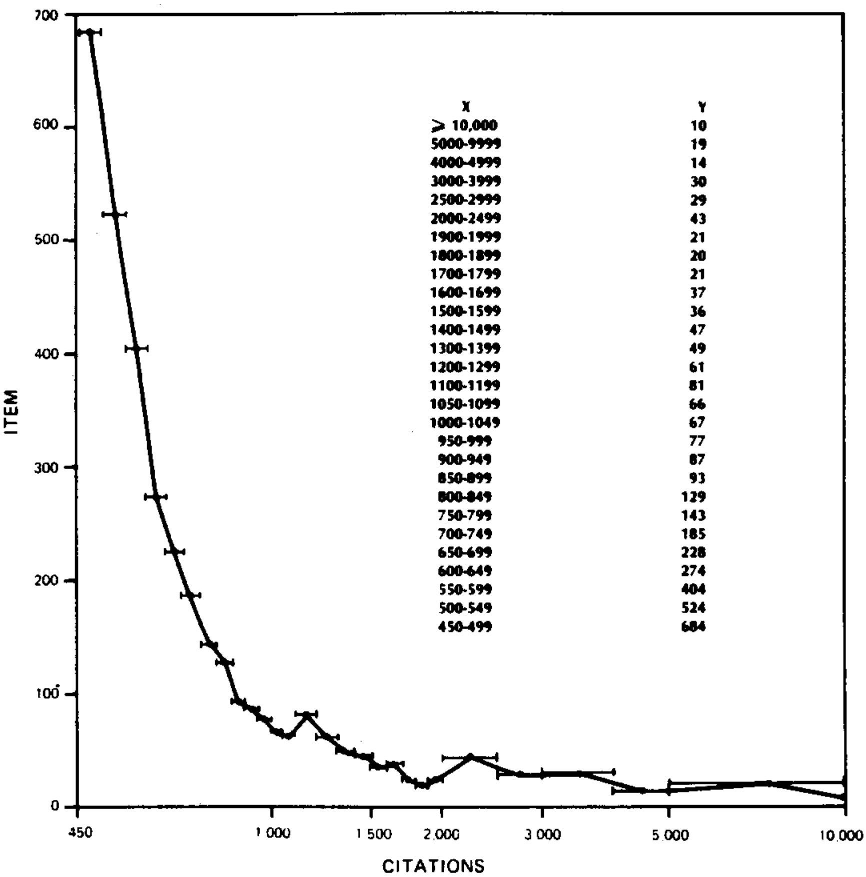
Unfortunately, no one has yet done an accurate census of the scientific literature. I estimate that the total number of published articles and books was about 30 million as of 1984. Therefore about one-sixth of the total literature may have been cited at least once that year. There are about 26 million unique items cited in our *SCI* file (covering 1955 to date). Most of this literature has been published very recently. The scientific literature is probably doubling every 15 years—almost 4 percent per year. So our database covers a very large percentage of the total extant scientific literature. To paraphrase Derek deSolla Price, 90 percent of the literature ever published was published in the past 20 years (3).

Citation frequency data are very interesting and very important. Henry Small, director of our Research and Development Department, uses such data, which ISI compiles routinely, to take a "sample" of literature each year and to identify the most active research fronts (4). Let me explain what we mean by a research front.

Citation frequency data follow a characteristic hyperbolic distribution (see Fig. 1). Many papers are never cited or are cited only once. (In the universe of "papers," ISI includes *all* published items, many of which, such as letters or preliminary communications, may not ever warrant a citation.) I would say, from what I have seen lately, that about 25 percent of the literature is never cited. At the other end of the citation spectrum are the highly cited papers. And an important aspect of these frequently cited papers is how often pairs of them are *co*-cited in the bibliographies of other papers.

Every year we begin our co-citation analysis by creating a file of 25,000 to 100,000 of the most-cited papers for that year. These are called "core papers." A core paper in one year will not necessarily turn up in the core in subsequent years and should not be confused with a "classic," which may or may not turn up as a core paper, depending upon its age and utility.

A cluster of core papers is the profile for defining a research front. Of the approx-



X	Y
≥ 10,000	10
5000-9999	19
4000-4999	14
3000-3999	30
2500-2999	29
2000-2499	43
1900-1999	21
1800-1899	20
1700-1799	21
1600-1699	37
1500-1599	36
1400-1499	47
1300-1399	49
1200-1299	61
1100-1199	81
1050-1099	66
1000-1049	67
950-999	77
900-949	87
850-899	93
800-849	129
750-799	143
700-749	185
650-699	228
600-649	274
550-599	404
500-549	524
450-499	684

FIGURE 1. Citation distribution for all items cited at least 450 times in the *SCI*®, 1961–1982. Items were tabulated for ranges of citations. The midpoint of each citation range is plotted. The horizontal bars through the points indicate the actual ranges. The table of values shows the actual ranges of citation values (x) and of number of items (y). (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

imately 565,000 items now indexed per year by ISI, a substantial number are not assigned to any of the 10,000 initial clusters we create. We use an arbitrary cutoff point that excludes a large number of "miscellaneous" small research topics each year.

The objective of "clustering" is to obtain a manageable number of useful subject-matter categories. Once identified, we then arrange them into a hierarchical scheme of five levels. At the lowest level, we are close enough to the center of the research fronts that we can see the individual core papers that identify them. Each level above that expands to include more and more papers citing the core, until, at the highest level, we have the "global map" of science, showing the broad areas of scientific endeavor for each year. The purpose of the year-to-year mapping exercise is to observe changing patterns in research and to enable scientists to see what direction their fields are taking.

To create an initial, or C1, cluster, we select a small percentage of the annual file of highly cited papers, namely those that have been significantly *co*-cited in the cur-

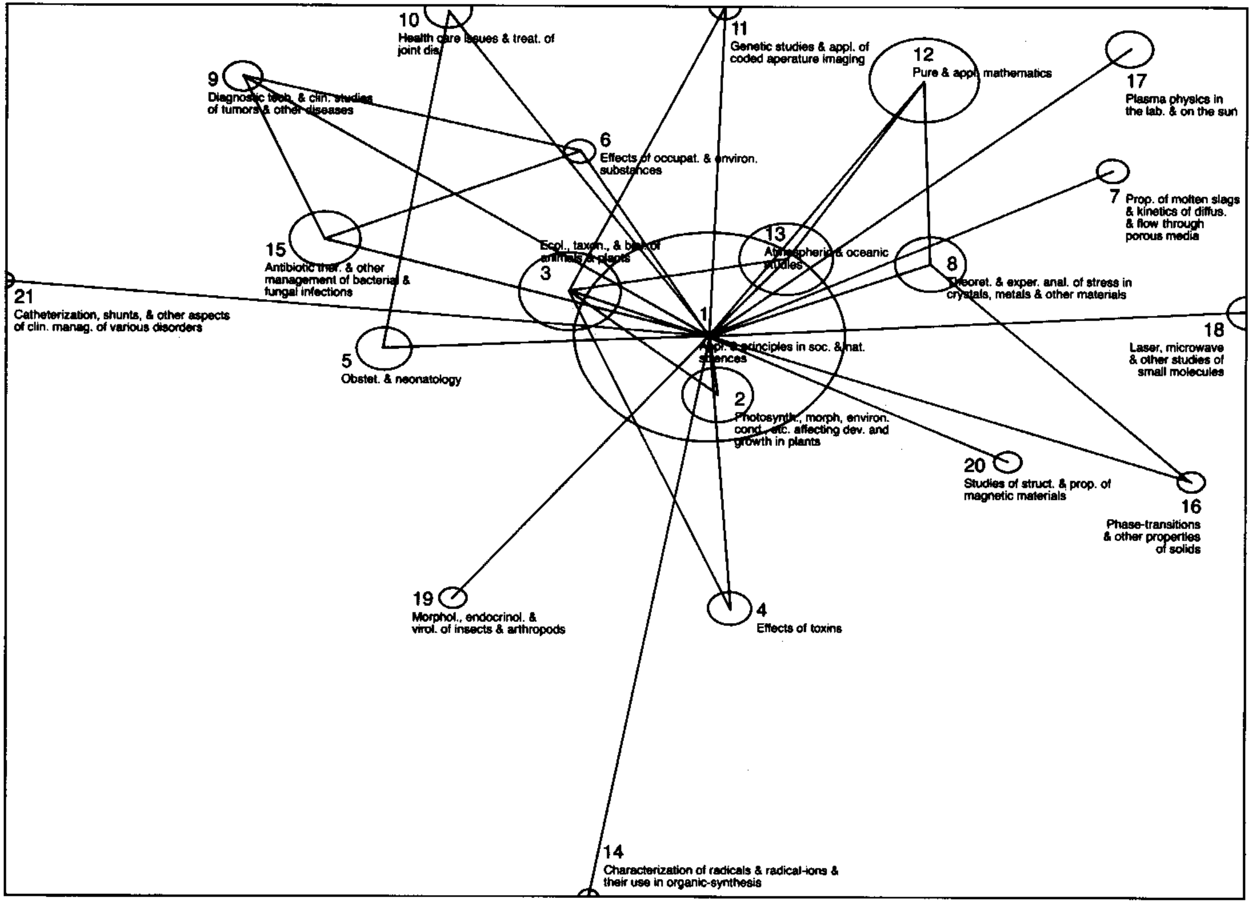


FIGURE 2. The 1984 global map of science. C5 cluster showing the various areas of current scientific research and their interrelationships. (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

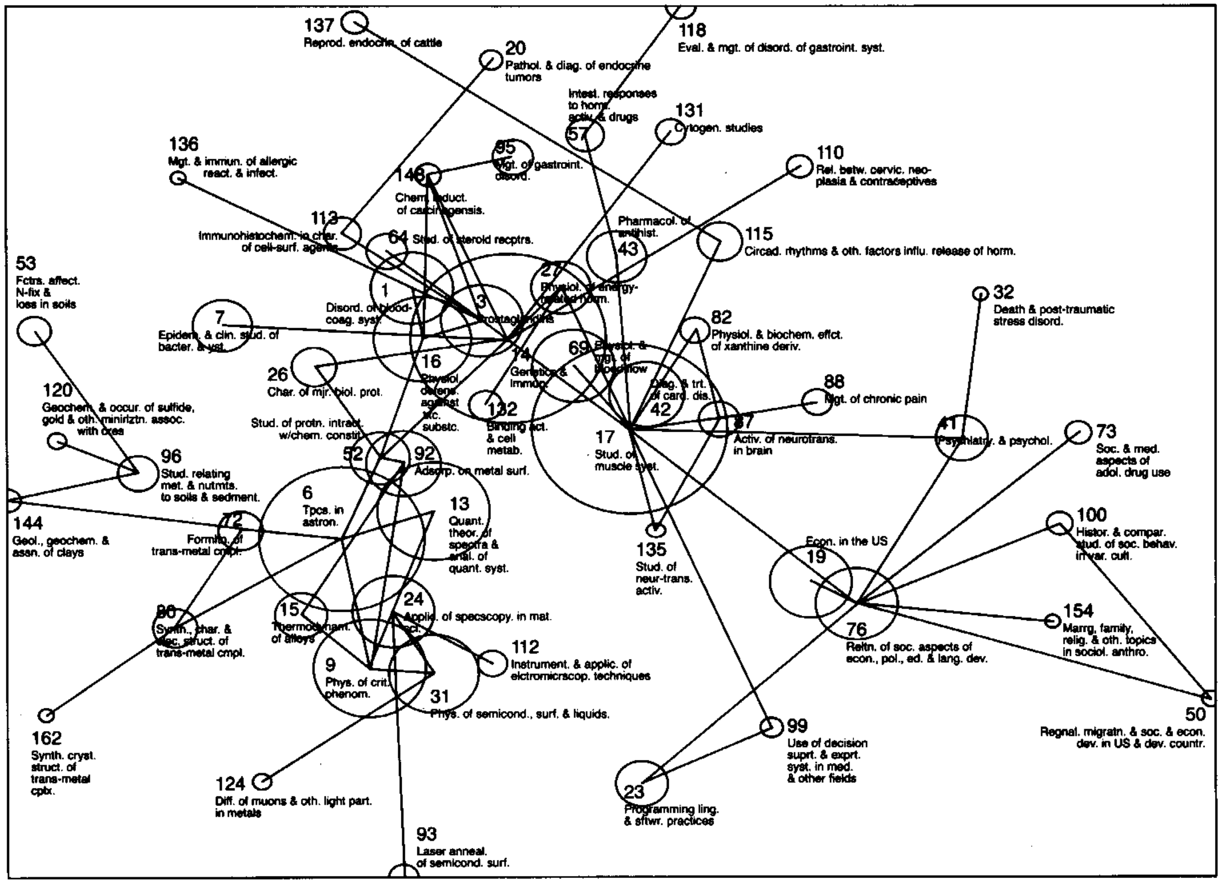


FIGURE 3. Map of C4 cluster #84-0001 showing the major research areas in the social and natural sciences. (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

rent literature. The two parameters that identify papers for a cluster are *citation frequency* and *co-citation strength*, which is calculated according to the formula:

$$\frac{\text{co-citations of A and B}}{(\text{total citations of A} + \text{B}) - (\text{co-citations of A} + \text{B})}\ .$$

If we set these parameters too high, no papers would qualify; if we set them too low, a substantial portion of the file would be selected and form one large, rather uninformative cluster called "science." Therefore, heuristics determine how the parameters are arbitrarily set to produce a cluster of "reasonable" size. For example, we can select papers cited 20 times or more with a co-citation strength of 24 percent. In our multidimensionally scaled maps, the inverses of the co-citation strengths are the distances between papers or clusters on the map.

By using the co-citation technique, we can create an initial cluster of papers that connect in interesting ways. If papers are co-cited frequently, they usually concern connected subjects, i.e., there must be some "topic" that brings them together, but the individual core papers do not necessarily treat the same subjects. Suppose we change the two parameters to 17 or more citations and 22 percent co-citation strength. By doing so, more core papers get into the cluster. This may result in a group of 30 to 40 core documents on a given subject. If some minimum information, such as the year of publication and author's address, is added for each paper, a map of the "invisible college" for a given subject will emerge.

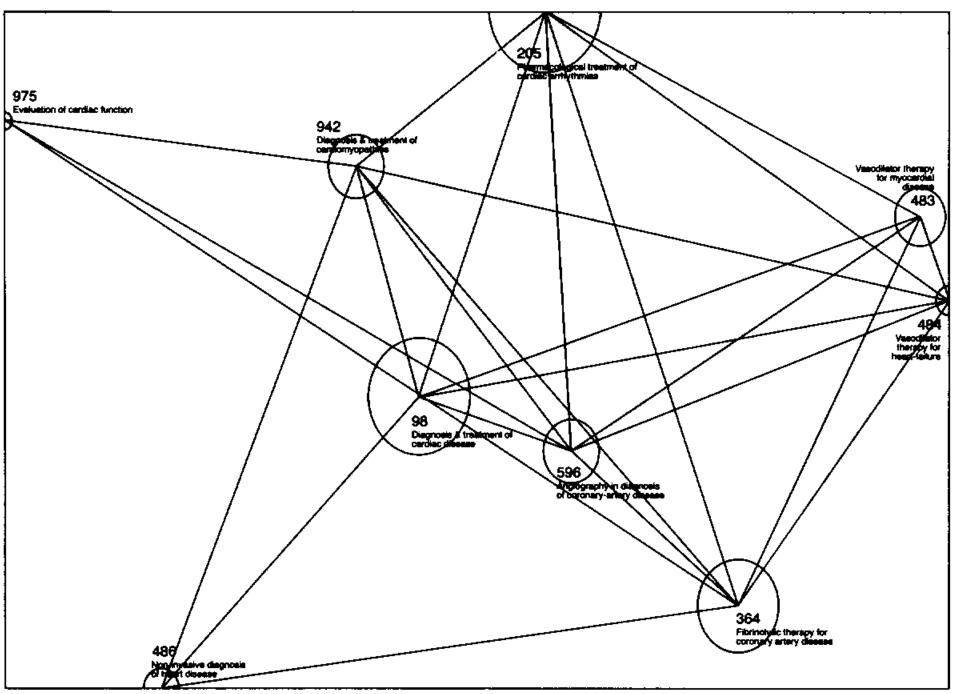


FIGURE 4. Map of C3 cluster #84-0042 showing the major research areas in the diagnosis and treatment of cardiac disease. (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

The process is identical for any subject the researcher may choose. We could expand this clustering process by progressively lowering the selection thresholds until literally an entire discipline is included. Instead we create precise initial (C1-level) clusters and then cluster the clusters. The C2, C3, and C4 levels show intermediate maps of scientific disciplines. The C5 level is the "global" map of science.

Maps are generally created for one year, but they can be created for varying time spans. Depending on how fast the literature changes and how much is published, maps for other time spans can be very interesting. In a fast-growing field such as AIDS research, the literature is expanding so rapidly that the map changes almost every month. A problem we face when making higher-level maps is this continuous shifting among the relationships and interests of disciplines. For example, since no one does old-fashioned "chemistry" anymore, it is not a useful discipline name on a higher-level map. Accurate labels for thousands of lower-level clusters are also problematical.

To get an idea of the varying scales on which these maps are made, I want you to imagine that we're on the moon looking down at the Earth. The Earth here represents the world of science. If you look at it with just a simple telescope, you will see the very vaguely defined "continents" of the world of research (see Fig. 2).

The C5-level ("global") map in Fig. 2 shows not only the interrelationships but also the relative sizes of the disciplines or clusters and their conceptual (semantic) distances from each other. One of the interesting results we found from comparing our

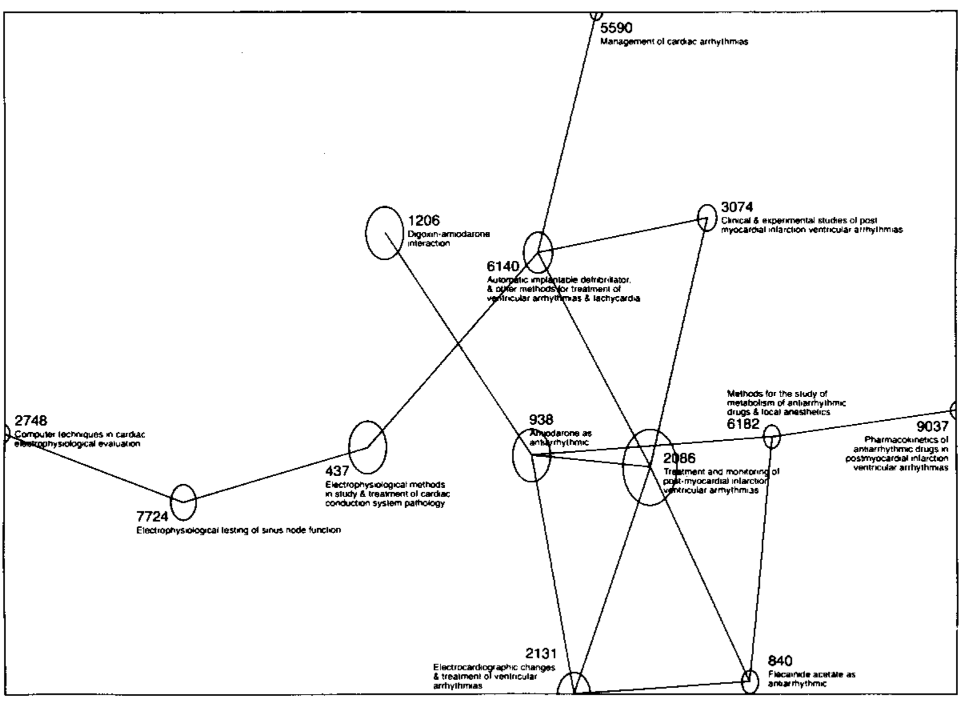


FIGURE 5. Map of C2 cluster #84-0205 showing the major research areas in the treatment of cardiac arrhythmias. (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

1984 C5 map with our 1983 C5 map was the convergence of the social and natural sciences due to the increasing amount of research in the neurosciences that is virtually dragging the whole "continent" of the social sciences into the "continent" of the natural sciences.

Focusing with a better telescope on the "continent" (C4 cluster) named "Social and Natural Sciences" (Fig. 3), we see different countries or regions (C3 clusters) such as "Diagnosis and Treatment of Cardiac Disease (cluster 42)," "Thermodynamics of Alloys (cluster 15)," and so on. Zooming in on cluster 42 (Fig. 4), we can pick out C2 cluster 205, "Treatment of Cardiac Arrhythmias" (Fig. 5). Here you can see all of the topics that lead into and out of this research area.

Finally, we look right at the center of this research area—at C1 cluster 6140 (Fig. 6)—"Automatic Implantable Defibrillators and Other Methods for Treatment of Ventricular Arrhythmias and Ventricular Tachycardia" to see what that specific research front contains. Thus, the maps go all the way from the global world of science down to this locality of core papers and scientists specifically concerned with implantable defibrillator research.

Another way of looking at the biomedical engineering field is by examining citation patterns among core biomedical engineering journals. Table 1 lists the core biomedical engineering journals covered in the *SCI* and shows the number of articles each journal published in 1984. The table also shows the number of times each journal cited the other core biomedical engineering journals in 1984. These 19 biomedical engineer-

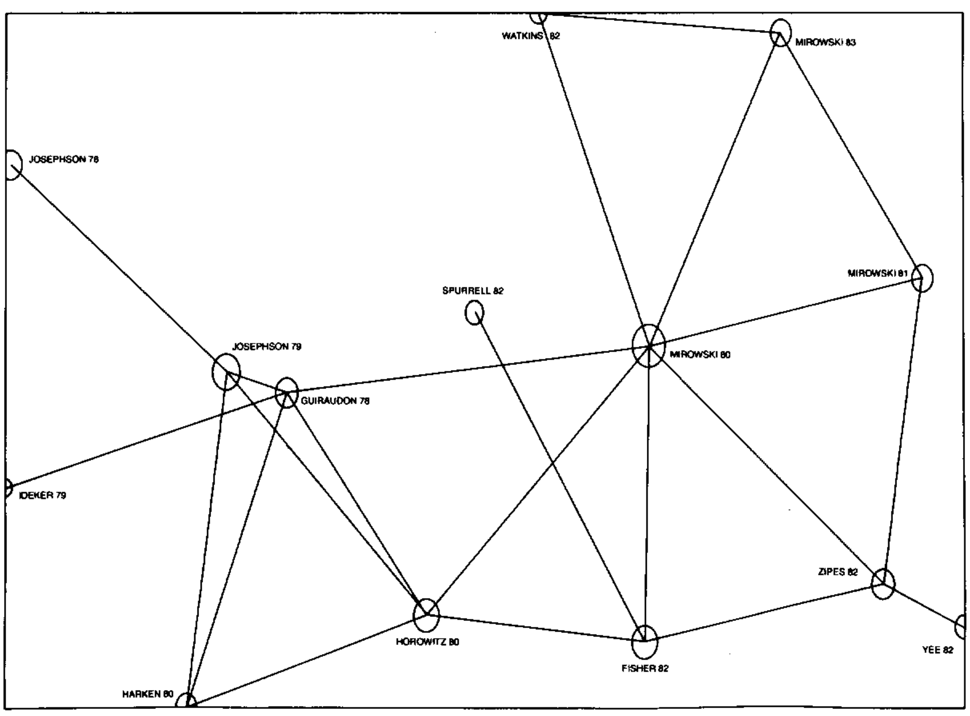


FIGURE 6. Map of C1 cluster #84-6140 showing the principal authors engaged in research on methods of treating ventricular arrhythmias and tachycardia, including implantable defibrillators. (Copyright 1985 by the Institute for Scientific Information, Inc.® Reprinted with permission.)

TABLE 1. Names of the biomedical engineering core journals, the total number of articles published in them in 1984, and the number of citations to biomedical engineering core journals originating in them in 1984.

Biomedical Engineering Journals Indexed in the *SCI*	Total Number of Articles Published	Number of Citations to Biomedical Engineering Core Journals
Annals of Biomedical Engineering	18	22
Artificial Organs	61	60
Biomaterials	65	162
Biomedizinische Technik	45	32
Clinical Physics and Physiological Measurement	30	5
Computers in Biology and Medicine	48	54
Computers and Biomedical Research	45	35
Computer Programs in Biomedicine	24	0
CRC Critical Reviews in Biomedical Engineering	21	30
IEEE Transactions on Biomedical Engineering	133	417
International Journal of Artificial Organs	61	79
Journal of Biomechanical Engineering—Transactions of the ASME	59	249
Journal of Biomechanics	93	370
Journal of Biomedical Engineering	59	63
Journal of Biomedical Materials Research	91	377
Journal of Medical Engineering and Technology	31	12
Medical and Biological Engineering and Computing	112	165
PACE—Pacing and Clinical Electrophysiology	187	432
Ultrasonic Imaging	25	71
	1,208	2,635

ing journals published 1,208 articles in 1984, which is 0.2 percent of the approximately 565,000 published articles ISI indexed that year. Since the biomedical engineering research fronts contain far more than 1,208 papers, we know that many papers in this field are published outside its core journals.

We can look at this publication and citation activity more clearly if we compare the statistics for the two sets of research fronts listed in Tables 2 and 3. Each table contains the top five 1983 biomedical engineering research fronts based on two different kinds of rankings. In both tables, the column marked "number of items in research front" gives the total population of papers published on the research subject noted. The column marked "number of items in core journals" gives the number of papers in the research front that were published in the core biomedical engineering journals in Table 1. "Percent in core journals" is the ratio of number of items in the research front to the number in the core journals.

Table 2 lists the five 1983 research fronts with the highest number of citing (published) articles from the core biomedical engineering journals. The top-ranked front (#83-0533) contained more articles—63—from the biomedical engineering journal set than did any other 1983 research front. This type of ranking identifies the research

TABLE 2. The 1983 biomedical engineering research fronts (C1 clusters) with the highest number of citing articles from the core biomedical engineering journals.

Rank	No. of Items in Research Front	No. of Items in Core Journals	% Items in Core Journals	Cluster Name
1	296	63	21.28	Characteristics and Mechanisms of Cardiac Modes Such as Dual-Chamber and Atrial Synchronized Ventricular Pacing, and Regulation of Q-T Intervals (Cluster #83-533)
2	673	40	5.94	Pulmonary Edema and the Role of Neutrophils in Lung Injury and the Adult Respiratory-Distress Syndrome: Effect of Peritoneal Dialysis and Hemodialysis (Cluster #83-695)
3	209	40	19.14	Cellulose to Cellulase Hydrolysis, Ethanol Fermentation, Beta-Glucosidase Activity, and Other Aspects of Enzymatic Hydrolysis (Cluster #83-2513)
4	167	35	20.96	Arthroplasty, Total Replacement, External Fixation, and Other Biomechanical Treatments for Various Bone Disorders (Cluster #83-3437)
5	450	26	5.78	Methanogenic, Thermoacidophilic Archaebacteria in the Anaerobic Treatment of Soluble Wastes in Reactors: Bacterial Growth in Cultures and Rumen (Cluster #83-1342)

fronts covered heavily in biomedical engineering journals. As you can see by comparing the size of the total citing literature for these fronts with the number of biomedical engineering journal citations, these are multidisciplinary research fronts in which many papers are published in other than biomedical engineering journals. This illustrates that our tabulations are quite diverse and multidisciplinary in nature.

Table 3 lists the five 1983 research fronts with the highest percentage of participation by articles from the core biomedical engineering journals. Of the 14 papers citing the core literature of the top-ranked research front, nine—or 64.29 percent—were from the journals listed in Table 1. This ranking emphasizes those research fronts not as heavily cited outside the biomedical engineering journals. The highest-ranking fronts in this set are not very multidisciplinary in their impact as yet. They tend to be defined within the scope of biomedical engineering alone.

An interesting evolution has taken place recently in the field of biomedical engineering that illustrates how small topics grow into research fronts and then begin to define disciplines. When we attempted to choose the C1, C2, C3, and C4 examples discussed above from our 1983 clusters, we found that no C1 biomedical engineering clusters carried through beyond the C3 level. That is, these "local regions" of

TABLE 3. The 1983 biomedical engineering research fronts (C1 clusters) with the highest percentage of participation by articles from the core biomedical engineering journals.

Rank	No. of Items in Research Front	No. of Items in Core Journals	% Items in Core Journals	Cluster Name
1	14	9	64.29	Ankle and Knee-Joint Mechanical Forces During Acceleration and Various Cadences (Cluster #83-8909)
2	8	4	50.00	Mycelial Growth, Use of Microbeads, and Other Aspects of the Control of Penicillin Fermentation (Cluster #83-8663)
3	14	6	42.86	Assessment of Arterial Flow and Pulse-Wave Impedence in Post-Surgical Cases (Cluster #83-8239)
4	19	8	42.11	Doppler Signals and Other Non-Invasive Methods of Blood-Flow Determination (Cluster #83-6523)
5	19	7	36.84	Biomedical Applications of Heparinized Surfaces (Cluster #83-7311)

research interest had not yet coalesced with any of the C4 "continents" that make up the global C5 map. We recently completed our 1984 data analysis, however, and found that in this data set, biomedical engineering C1 clusters now carry through to the C5 level, and we have chosen one of them as the basis of our example here. This is an excellent illustration of the emergence and maturation of a scientific discipline as revealed by mapping.

At ISI we are now taking this concept of mapping one step further to create an atlas of science on a topic-by-topic basis. The collection of these atlases will constitute an encyclopedia of science. Each topical volume includes the key research fronts for that topic displayed as individual maps. For the core papers in each map we tabulate the key literature citing into that core. This list of key citing documents is determined by using a measure we call relevance weight: the number of core documents a paper cites. The more core papers cited by a paper, the higher its relevance weight. We can use the same core to prepare lists of citing papers for subsequent years, but their relevance weights may change over time. The speed with which the change takes place says something about how that core topic is changing.

An important aspect of the atlases is that, associated with each map, core list, and list of key citing documents, there is a prospective review. These reviews are written by experts and succinctly summarize each topic. Unlike ordinary review papers and annual reviews, which do not cover every scientific field and in fact leave significant gaps, our procedure is systematic. We can identify all subjects in science needing a review. We will also be able to show the trends in scientific research, which fields are emerging, which ones are converging, and so on. This is what mapping is all about.

REFERENCES

1. Garfield, E. The mystery of the transposed journal lists—wherein Bradford's law of scattering is generalized according to Garfield's law of concentration. In *Essays of an Information Scientist*, Vol. 1. Philadelphia: ISI Press, 1977, pp. 222–223.
2. Institute for Scientific Information. *SCI-Science Citation Index, 1984. Annual.* Philadelphia: ISI Press, 1985, 14 vols.
3. Price, D.J.D. *Little Science, Big Science*. New York: Columbia University Press, 1963.
4. Small, H. and B.C. Griffith. The structure of scientific literatures. I: Identifying and graphing specialties. *Sci. Stud.* 4:17–40, 1974.