

## The Geography of Science: Disciplinary and National Mappings

Henry Small and Eugene Garfield

Maps of research in the sciences and social sciences for the year 1983 were generated from the *Science Citation Index*<sup>®</sup>/*Social Sciences Citation Index*<sup>®</sup> database using a single-link clustering algorithm. The combined use of fractional citation counting and co-citation frequency thresholds ensured a fair representation of smaller areas. The resulting clusters of co-cited publications were aggregated through iterative computer runs at five levels, C1 (local) to C5 (global). Multidimensional scaling at each level allowed the matrix of linked publications to be represented geometrically as maps. Each research front or node appears on the maps in proportion to the population (size) of its citing publications. Analyses of the distribution of research fronts by disciplines and nations are demonstrated, as well as measures of expectation and immediacy for research in genetic engineering and, in particular, on studies of the beta-thalassemia gene. A walking tour through the maps at various levels coupled with content analyses of linkages reveal that, in general, large central areas represent basic research whereas smaller peripheral specialties represent applied research. Although in its infancy, the construction of such maps of the sciences (and comparisons of maps over time) holds the potential for a new kind of history of science.

Each year ISI<sup>®</sup> undertakes an analysis of a database which is derived from a combination of the *Science Citation Index*<sup>®</sup> (*SCI*<sup>®</sup>) and the *Social Sciences Citation Index*<sup>®</sup> (*SSCI*<sup>®</sup>). The purpose of this analysis is to create what we call maps of science which show the topography of science at various levels of aggregation. These range in scale from a global perspective down to the level of the individual investigator's key papers. It is fair to say that the science and technology for creating such maps is in its infancy. Earlier work by Garfield, Sher, and Torpie<sup>1</sup> on historiographs and Price on networks of scientific papers<sup>2</sup> showed the longitudinal structure of science by exploiting citation links across time. More recent co-citation methods emphasize the cross-sectional structure of science by using co-occurrence of references<sup>3</sup> at a specific point in time, and hence suggest the analogy to geographic maps.

The notion that science can be mapped was first clearly stated by Derek Price<sup>4</sup> during the 1960s, though we are sure that a thorough search of the literature of information science, history, sociology and philosophy of science would reveal significant precursors. For example, in 1948 Samuel Bradford,<sup>5</sup> the famous British bibliographer, wrote:

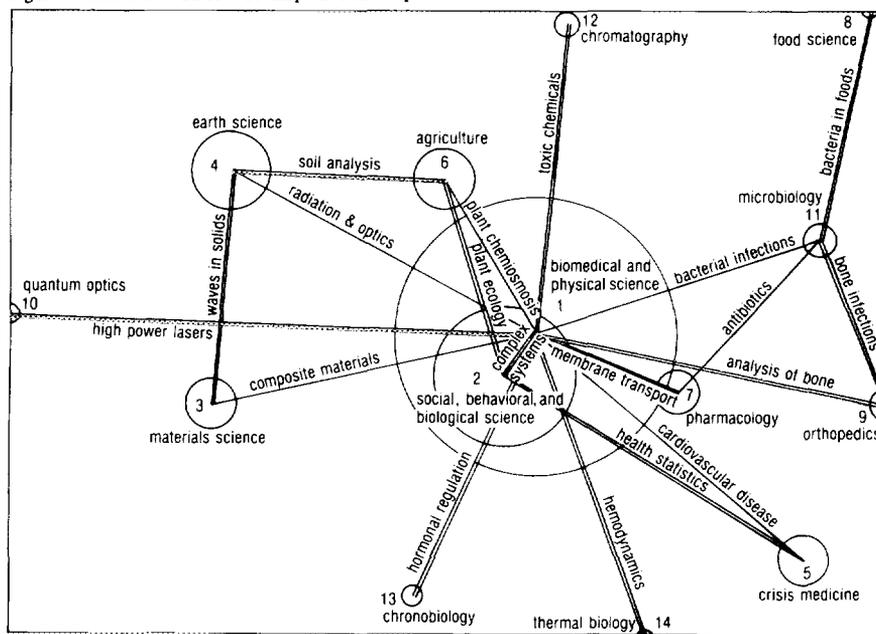
"...if  $x$  represents any class, as men, its individuals will have many different qualities. We may separate those individuals, which are distinguishable from one another as having different qualities, in sub-classes. The symbols of all sub-classes will have only the possible numerical values 0 and 1. So let us draw straight lines, of unit length, from a convenient point, to repre-

sent these symbols. The lines will be distinguished by direction, and, if we like to use a space of three dimensions, the unit lines will terminate in points upon the surface of a unit sphere. The aggregate of points will represent the class of beings, men. Similarly, let more lines be drawn to represent all things we wish to talk about. The aggregate of points, where all these lines end on the surface of the sphere, represents the universe of discourse.... And so we get a picture of the universe of discourse as a globe, on which are scattered, in promiscuous confusion, the mutually related, separate things we see or think about."

Bradford is certainly one of the precursors of our present effort, but the idea that knowledge could be represented spatially is implicit in many early statements. The notion of 'disciplines' or 'fields' carries with it spatial connotations, which were made more explicit by Vannevar Bush's metaphor of the scientific 'frontier.'<sup>6</sup> In social science, the concept of mental maps, which are subjective versions of geographic maps, has been important to some fields such as human geography.<sup>7</sup> In the sociology of science the concept of the invisible college suggests that communication between scientists can give rise to social groups.<sup>8</sup> But it is in the specialized branch of information science called bibliometrics that the idea of mapping science was finally realized.

Citation indexes, showing millions of interconnections among hundreds of thousands of scientific articles and books annually, seem ideally suited for deriving natural maps of the scientific landscape. No other database is as comprehensive either in disciplinary or longi-

Figure 1. 1983 *SCI*<sup>®</sup>/*SSCI*<sup>®</sup> C5 map: Global map.



tudinal scope as the *Science Citation Index* and the more recent *Social Sciences Citation Index*. In this brief paper it will be possible to give only a glimpse of ISI's annual mapping exercise, its results, and some practical applications.

In utilizing a combination of the *SCI* with the *SSCI* we attempted for the first time in 1983 to observe connections between the natural and social sciences. This mixing of disciplines also suggests the first problem we had to overcome, namely, normalizing for varying citation rates across different fields. The first step in the mapping process is to set a citation frequency threshold and select only the most cited documents for processing through a clustering algorithm. Since citation rates differ across fields of science, applying a simple integer threshold biases the selection to high citing fields—biomedical research eliminates less citing fields such as mathematics. The method we came up with, while perhaps not ideal, gives quite satisfactory results. It is called fractional citation counting and amounts to assigning to each published paper or citing item one unit of strength to be divided equally among all its references.<sup>9</sup> We can measure the effectiveness of this procedure by matching the rough disciplinary distribution of highly cited documents selected

by it to the distribution of source papers appearing in the index. If the percentages are comparable, then we are sampling the field in proportion to its representation in the database. For example, we know that about 14% of the source items in the database are from the social sciences. The number of cited documents in the social sciences selected by the fractional threshold was about 12%, indicating proportional coverage of that field.

Other strategies are used to ensure that we are obtaining a fair representation of large and small research areas. The cluster analysis is based on the co-citation frequency, which is the number of times two documents are cited together by current papers. This association measure is normalized by dividing by the square root of the product of the citation frequencies of the co-cited documents, a formula widely used in information science studies.<sup>10</sup> The clustering algorithm is called the single-link method and has the advantage of simplicity in applications involving large files but the disadvantage that its clusters are sometimes highly chained. To limit the amount of chaining we use a cutoff in cluster size of 60 cited documents and a variable co-citation level to increase the size of small clusters. Since the limit of 60 in cluster size means that some areas will be arbitrarily cut

Table 1  
 Statistics on iterative clustering of clusters: 1983 *SCI*<sup>®</sup>/*SSCI*<sup>®</sup>

	Iteration:				
	C1	C2	C3	C4	C5
1. fractional citation threshold (drop items cited $\leq 4$ times)	1.5	0	0	0	0
2. cited items selected	72 077	9420	1386	171	14
3. citations to cited items	1 155 257	517 354	329 540	219 478	179 117
4. mean citations per item	16.0	54.9	237.8	1283.5	12 794.1
5. distinct co-cited pairs	1 789 036	196 380	32 599	2755	62
6. percent connected	0.069%	0.443%	3.40%	18.9%	68.1%
7. co-citation threshold	0.17+	0.017+	0+	0+	0+
8. level increment	0.03	0.01	0.005	0.001	0
9. max cluster size	60	60	60	60	60
10. clusters	9420	1386	171	14	1
11. cited items in clusters	50 994	6018	729	111	14
12. co-cited pairs	63 111	5279	606	116	62
13. mean items per cluster	5.4	4.34	4.26	7.93	14
14. items in largest cluster	60	54	50	52	14
15. % clusters with two cited items	41.6%	49.3%	50.9%	42.8%	0%
16. % cited items clustered	70.7%	63.9%	52.6%	64.9%	100%

into pieces, we put the pieces back together by successive iterations of clustering.<sup>11</sup> That is, in the first step we cluster cited documents by co-citation. Then, in a second iteration we cluster clusters derived from the first step. We continue clustering the clusters of the previous step until we obtain a single mega-cluster which includes as many of the previous groupings as is possible. The structure of this final grouping corresponds to our global map of science (Figure 1). Later on we will discuss the structure of this map in detail.

In order to geometrically display what is really only a matrix of objects linked together by varying degrees we use another technique called multidimensional scaling.<sup>12</sup> Imagine taking a map of the United States and constructing a table showing the distances between all major cities. Our problem is the reverse. We have the table of distances (or actually degrees of closeness) but lack the map embodying those distances. This is what the scaling technique provides in, of course, an approximation depending on the number of dimensions required. In our case, we have used only two-dimensional representations, but there is no reason why the maps could not be three-dimensional.

The statistics on iterative clustering to five levels can be seen in Table 1 for the 1983 combined *SCI* and *SSCI*. We initially drop all items cited fewer than 5 times which provides a low cutoff for citedness and prevents the introduction of random noise. The fractional threshold of 1.5 translates into a variable inte-

ger citation cutoff. For example, only about one-half of all items cited 20 or more times are included with a fractional cutoff of 1.5, while about 10% of items cited 5 times are included. Hence we retrieve varying percents for each integer citation level up to about 40 times cited where 100% of items are retrieved. These varying samples help compensate for differences in citation rates across fields.

Of the many statistics generated from clustering, perhaps the essential ones are the number of clusters generated at the first iteration (called C1), namely, 9,420 for 1983, and number of cited documents contained therein, namely, 50,994. This calculates to a mean of 5.4 cited items per cluster, with a maximum cluster size of 60 items. As we move through each iteration, C1 through C5, the file is increasingly aggregated, and fewer and fewer clusters form until only one is formed at C5. When processing is completed, a nested hierarchy of clusters is generated which is five levels deep. This means that a point on the highest level map corresponds to a map of points on the next lower level, and so on down for five levels. The number of iterations or levels required to reduce the file to a single grouping of 60 or fewer macro-clusters depends on the number of cited documents selected by the fractional citation threshold which are input to the clustering process.

One of the deficiencies of our present method which we are working to improve is the inability of the iterative clustering scheme

Table 2  
1983 *SCI*<sup>®</sup>/*SSCI*<sup>®</sup> clusters: disciplinary distribution

	C1		C2		C3		C4	
	number of clusters	percent						
biomedicine and biochemistry	3625	38.5	506	36.5	59	34.5	7	50.0
agriculture, agronomy, botany, entomology, ecology, animal and food science	690	7.3	126	9.0	20	11.7	2	14.3
physics and materials science and engineering	1696	18.0	295	21.3	38	22.2	2	14.3
chemistry	1266	13.4	166	12.0	16	9.4	1	7.1
social and behavioral sciences and psychiatry	1099	11.7	152	11.0	16	9.4	1	7.1
mathematics and computer science	576	6.1	87	6.2	13	7.6	0	0
geosciences	468	5.0	55	4.0	9	5.2	1	7.1
	9420	100.0%	1386	100.0%	171	100.0%	14	99.9%

to include all clusters obtained at a given level or iteration in the macro-clusters obtained at the next higher level. For example, as Table 1 shows, about 36% of the clusters at C1 are not included in C2 macro-clusters and hence remain isolates. The global map at C5 includes only about one-third of the C1 clusters. The creation of isolates is due to a combination of factors including the use of a minimum cluster size of two and a maximum cluster size limit of 60. While it may be impossible to entirely avoid creating isolates with the present methods, it should be possible to adjust the parameters to ensure that at least all of the largest clusters are included on the maps.

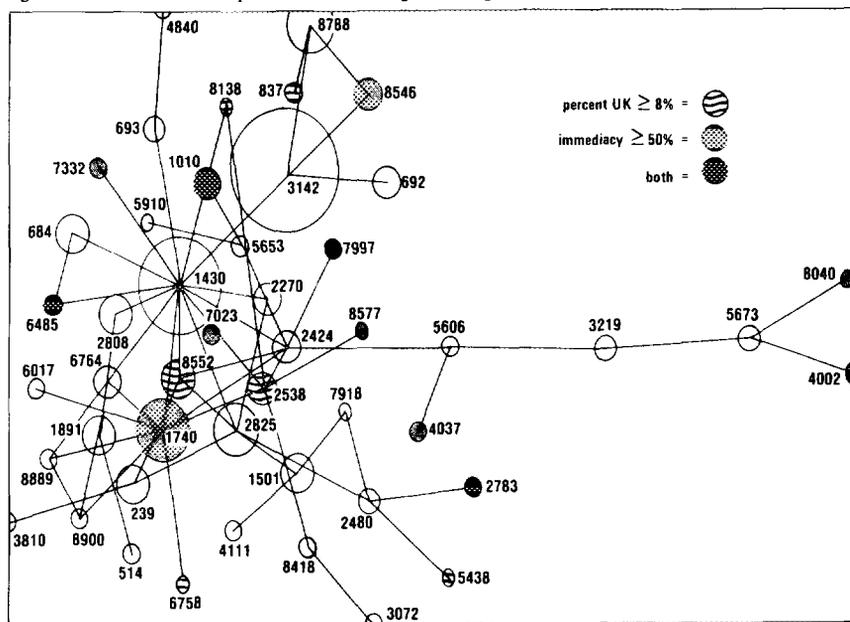
One of the most interesting results of our analysis is the distribution of clusters by disciplines, defined more or less in the traditional sense. We find, for example, 38.5% of clusters at C1 are on biomedical or biochemical topics, 7.3% on ecology and systematic biology, 18.0% in physics, 13.4% in chemistry, 11.7% in social and behavioral sciences, 6.1% in mathematics and computer science, and 5.0% in geosciences (Table 2). As we move up the hierarchy from C1, changes in the percentages reflect the varying tenden-

cies of fields to aggregate or to remain separate. For example, biomedical areas decline in their percent of clusters at C3 while ecological clusters increase, showing a greater tendency of the biomedical areas to aggregate.

The distribution of national effort across clusters is also easily determined. There are very few which are totally dominated by a single country. Cases of this type are small clusters usually on topics of particular national concern. However, patterns of national emphasis are easily discerned. Country participation in a cluster is gauged by comparing the number of papers from a country with the overall participation of that country in the file as a whole. Thus for any given cluster, it is possible to determine whether the country is participating less or greater than expected on the basis of a random distribution of a country's research effort.

The disciplinary and national data for clusters can be combined in the following way. If we are interested in how a country is concentrating its research effort, we can examine all clusters where that country's participation is at, or above, the expected level and look at their disciplinary distribution compared to

Figure 2. 1983 C2 cluster map, cluster no. 0145: genetic engineering.



KEY

239 gene mapping	2808 DNA methylation	6485 human gene cloning
514 histone variants	2825 cDNA cloning in E-coli	6758 oncogene products
684 DNA cloning	3072 tRNA gene transcription	6764 mammary tumor virus
692 DNA binding proteins	3142 protein detection	7023 gene cloning
693 cDNA clones	3219 genetics of immunoglobulin	7332 human genomic families
837 protein electrophoresis		7918 new gene elements
1010 beta-thalassemia gene	3810 DNA promoters	7997 tumor antigen gene
1430 cDNA cloning from mRNA	4002 T-cell regulation of B-cells	8040 genetics of antibodies
1501 E-coli cloning	4037 yeast gene polymorphism	8138 human globin genes
1740 oncogenes & cancer	4111 gene fusions	8418 RNA transcription
1891 chromatin DNA	4840 mRNA synthesis	8546 protein purification
2270 yeast genes	5438 DNA sequences	8552 nucleotide sequences
2424 globin genes	5606 histocompatibility genes	8577 ribosomal RNA
2480 transposable genes	5653 cDNA for beta globin	8788 protein synthesis & expression
2538 gene transcription	5673 T-cell idiotypes	8889 herpes virus gene
2783 hybrid dysgenesis	5910 mutagenesis	8900 gene expression
	6017 simian virus-40	

world science, as represented by the file as a whole. This was done recently for the United Kingdom and it was found consistently for all levels that the U.K. was below world averages in physics (-4%) and chemistry (-3%), while it was above world averages in biomedical research (+5%) and ecology (+2%) (see Table 3). Whether this view is shared by U.K. scientists or whether it reflects a conscious science policy in the U.K. remains to be seen.

Country concentration can be seen in individual cases by coding the cluster maps. The circle around each node on the map is proportional in area to the size of the cluster as

measured by number of citing papers. We then indicate by shading those areas where participation from a given country (in this case the U.K.) is greater than expected. For example, on the above map showing genetic engineering areas (Figure 2), research on the beta-thalassemia gene (#1010) is one of eleven areas in which the U.K. has a representation greater than expected (18% actual versus 8% expected). The expected rate is simply the percentage of papers contributed by that country to the whole file.

Another kind of coding is by immediacy of the cluster. Since each cluster or macro-clus-

Table 3  
1983 *SCT*<sup>®</sup>/*SSCT*<sup>®</sup> clusters: United Kingdom disciplinary distribution<sup>a</sup>

	C1		C2		C3	
	number of clusters	percent	number of clusters	percent	number of clusters	percent
biomedicine and biochemistry	1520	43.0 <sup>b</sup>	256	43.9 <sup>b</sup>	40	50.6 <sup>b</sup>
agriculture, agronomy, botany, entomology, ecology, animal and food science	348	9.8 <sup>b</sup>	60	10.3 <sup>b</sup>	12	15.2 <sup>b</sup>
physics and materials science and engineering	506	14.3 <sup>c</sup>	96	16.5 <sup>c</sup>	8	10.1 <sup>c</sup>
chemistry	366	10.3 <sup>c</sup>	54	9.3 <sup>c</sup>	5	6.3 <sup>c</sup>
social and behavioral sciences and psychiatry	377	10.7 <sup>c</sup>	59	10.1 <sup>c</sup>	6	7.6 <sup>c</sup>
mathematics and computer science	213	6.0 <sup>c</sup>	32	5.5 <sup>c</sup>	3	3.8 <sup>c</sup>
geosciences	208	5.9 <sup>b</sup>	26	4.5 <sup>b</sup>	5	6.3 <sup>b</sup>
	3538	100.0%	583	100.1%	79	99.9%

<sup>a</sup> counting clusters having greater than expected (8%) U.K. participation.

<sup>b</sup> greater than expected compared to world.

<sup>c</sup> less than expected compared to world.

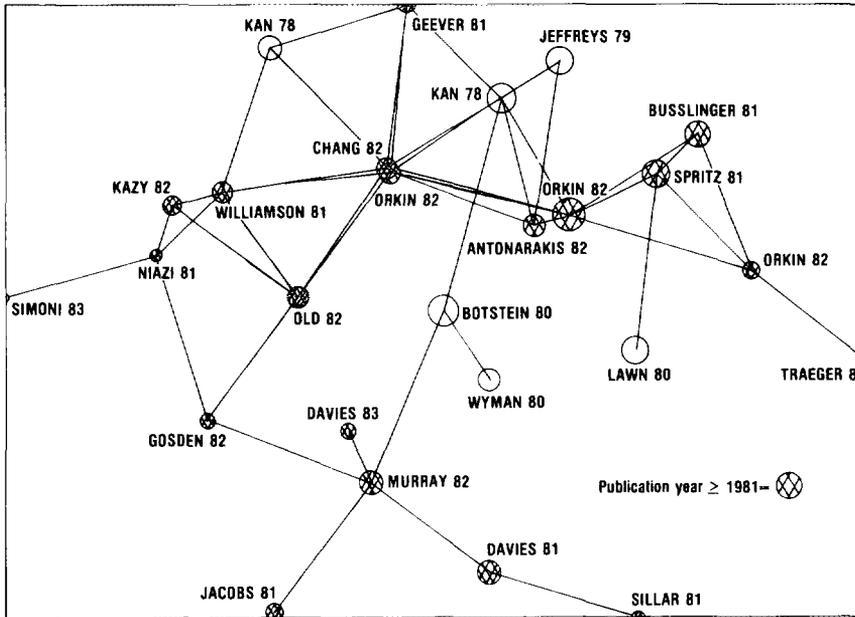
ter consists of a set of cited documents of varying age, an immediacy measure can be devised to reflect how recent this core literature is. The percentage of core documents published within the last three years is one such measure. On Figure 2 each area having an immediacy greater than or equal to 50% is indicated by a different shading of the circle. There are twelve such areas on the map, one of which is beta-thalassemia, with greater than expected U.K. participation. (Five areas have both high immediacy and high U.K. participation.) To show how immediacy is determined, the C1 map for beta-thalassemia is shown in Figure 3, with circles of varying size representing cited documents, scaled to reflect relative citation frequency and labeled with first author and publication year. The documents published in the last three years are shaded, and the number of such shaded circles divided by the total gives an immediacy of 73% (19/26).

Returning to the global map, we will now take an excursion starting with the C5 level of science to illustrate how the system works. From any node on the global map we can zoom down to the next lower level and observe its structure. Positioning at the large central circle representing biomedical and physical sciences, the map for this macro-cluster is displayed (Figure 4). We see three

major regions on this map: biomedical research on the left, chemistry in the middle, and physics and mathematics on the right. This 'in between' position for chemistry has been a consistent finding over the years we have performed this analysis. Focusing on the large circle labeled 'molecular genetics and immunology,' we move to the next lower level and see its map (Figure 5). Shown here is the genetic engineering area (#145) and a large immunotherapy area (#31) to its right, the site of work on AIDS and monoclonal antibodies. Other branches of this map concern the use of x-ray crystallography to study the structure of biological molecules (on the right), the biochemistry of connective tissues (on the left), and isolation and purification methods (at the bottom). Zooming in on the large genetic engineering area takes us up to the map shown in Figure 2, containing the beta-thalassemia cluster, Figure 3.

Moving up again to the global map, we can descend into the circle left and slightly below center representing the social and behavioral sciences, and systematic biology (Figure 6). The map for this region is clearly divided into a social/behavioral region on the left and a biology/ecology region on the right. Spanning these regions is work on multivariate statistics and classification. This somewhat tenuous link is the main bridge between the so-

Figure 3. 1983 C1 cluster map, cluster no. 1010: beta-thalassemia gene.



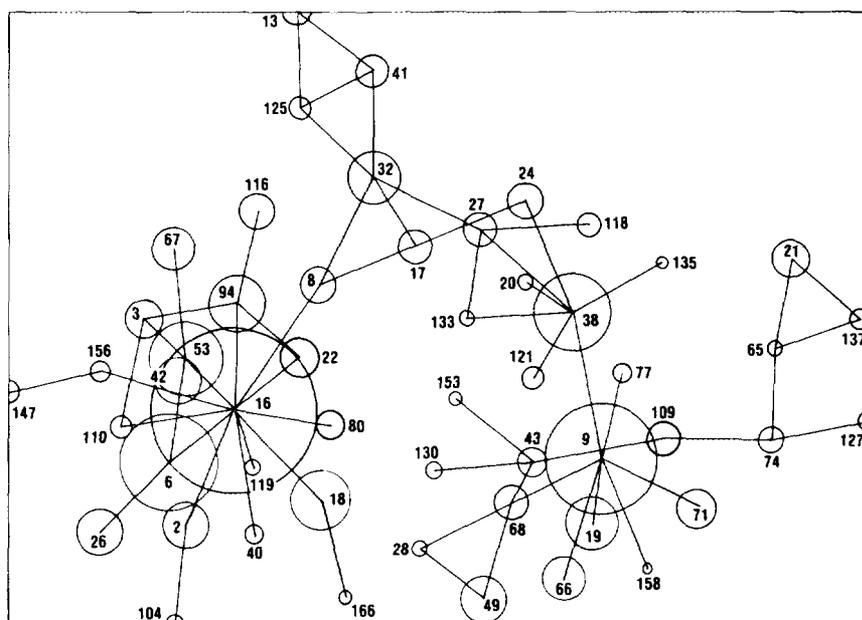
cial and biological sciences, at least in 1983. It is interesting to note that this conjunction of social and biological sciences is mirrored in the administrative structure of a major U.S. funding agency, the National Science Foundation.

The C5 map holds special interest because it is the most inclusive map in the hierarchy, containing roughly one-third of the C1 clusters within its 14 C4 macro-clusters. First we note that it is a combination of a few very large areas and several smaller ones. These size differentials reflect the essential hyperbolic nature of all bibliometric distributions, including the distributions of cluster size. Second we note that the pattern of links is predominantly center-periphery, with the large central area serving to link the various smaller ones around it. Very often the nature of these peripheral areas is more applied and less basic than the central areas, although this is not always the case. Examining the subject matter arrangement, we note that there is a rough division between physical and life-science areas on either side of a line drawn from the lower left to the upper right-hand corner. Only relative locations of objects on these maps are significant, however.

The scale and inclusiveness of areas on the global map vary widely. The two large central areas (#1 and #2) are multidisciplinary in character, encompassing biomedical, chemical, physical science, and mathematics in one case and social and behavioral sciences, systematic biology, and ecology in the other case. Most of the medium-sized areas (agriculture #6, earth science #4, materials science #3, pharmacology #7) correspond to familiar disciplinary groupings, with the exception of crisis medicine #5. This area might also be called emergency or trauma medicine including aspects of intensive care, and its emergence shows how an automatic classification method can create new categories which reflect contemporary concerns. The smallest areas are more akin to specialties which, for reasons we do not yet fully understand, have remained separate from the larger disciplinary groups, yet have links to them. Some of these small peripheral areas are applied science topics with connections to basic science.

The reasons for the links between the macro-clusters are as important as their specific contents. These links are the threads which hold the fabric of science together. We

Figure 4. 1983 C4 cluster map, cluster no. 0001: biomedical and physical science.



**KEY**

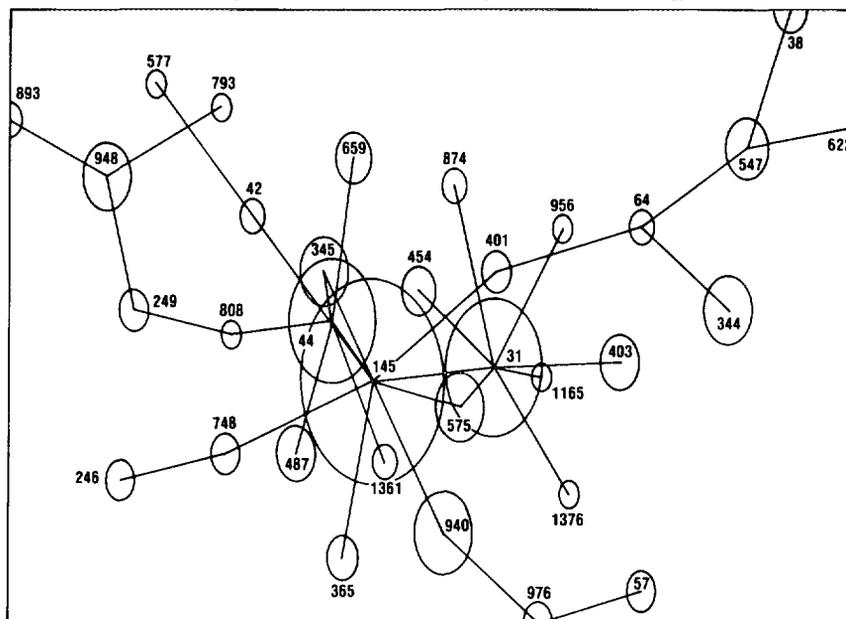
- |                                      |   |  |
|--------------------------------------|---|--|
| 2 cancer                             | 28 group theory                           | 109 control theory                           |
| 3 immune-complexes & autoimmunity    | 32 photochemistry                         | 110 immunoglobulin & secretory immune system |
| 6 neuroscience                       | 38 molecular orbital theory               | 116 pulmonary disease                        |
| 8 protein structure & dynamics       | 40 parasitic infections                   | 118 macrocyclic compounds                    |
| 9 physics, general                   | 41 polymers & glasses                     | 119 viral diseases                           |
| 13 thermodynamics                    | 42 cardiology: diagnosis                  | 121 natural products                         |
| 16 molecular genetics & immunology   | 43 Banach-spaces                          | 125 hydrodynamics & rheology                 |
| 17 electrochemistry                  | 49 algebraic topology                     | 127 numerical methods                        |
| 18 assays for carcinogens & mutagens | 53 cardiology: treatment                  | 130 uncertain dynamic systems                |
| 19 surface physics                   | 65 computational algorithms               | 133 C <sub>13</sub> NMR                      |
| 20 organometallic complexes          | 66 nuclear physics                        | 135 heteroaromatic compounds                 |
| 21 computer programming              | 68 quantum field theory                   | 137 graph theory algorithms                  |
| 22 coagulation & platelet disorders  | 71 critical phenomena                     | 147 metabolic enzymes                        |
| 24 transition metal complexes        | 74 information theory                     | 153 stochastic processes                     |
| 26 endocrinology                     | 77 fiber-optics                           | 156 liver metabolism                         |
| 27 reaction kinetics                 | 87 diet, metabolism & ion transport       | 158 electronic properties of solids          |
|                                      | 94 prostaglandins & inflammatory response | 166 lipids & toxicity of foods               |
|                                      | 104 allograft survival & transplantation  |  |

can analyze the nature of these links by sampling papers which cite core documents in the adjacent areas, and then examining the portion of their texts in which the specific references are made. The technique, known as citation context analysis,<sup>13</sup> has been carried out for all links shown on the global map (Figure 1) and the phrase best describing the nature of the link is indicated. In addition, certain links on the map have been 'doubled.'

These links define a minimal spanning tree through the network beginning with the food science node in the upper right. We will follow this path to organize our description of the kinds of links which tie together these large areas.

Food supplies are subject to bacterial contamination by various organisms, for example, *Salmonella*, and the link to microbiology is due to the use in food science of assays

Figure 5. 1983 C3 cluster map, cluster no. 0016: molecular genetics and immunology.



**KEY**

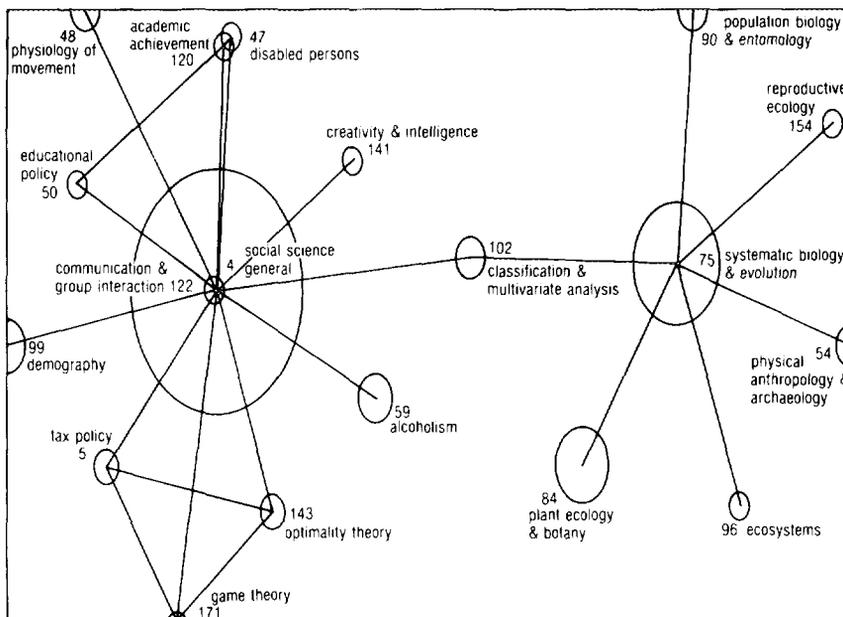
31 T & B cell subsets in immunotherapy	365 lipoproteins of bacteria	808 atherosclerosis
38 NMR of biological systems	401 DNA secondary structure & binding	874 anti-tumor activity of T & B cells
42 structure of RNAs	403 Epstein-Barr virus	893 glycosaminoglycans & proteoglycans
44 protein biosynthesis	454 bone marrow transplantation	940 isolation & purification of proteins
57 iron-overload disorders	487 tumor promoters	948 collagen & basement membranes
64 x-ray & NMR conformation analysis	547 NMR & protein structure	956 suppressor cell activity
145 genetic engineering	575 histocompatibility genes	976 automated analyses of amino-acids
246 receptor binding of hormones	577 ribosomal-RNA & proteins	1165 allograft-rejection & monoclonal antibodies
249 angiogenesis by vascular endothelial cells	622 nuclear-Overhauser-enhancement	1361 glycoprotein biosynthesis
344 x-ray crystallography of organometallic complexes	659 insulin receptors & diabetes	1376 mixed-lymphocyte reactivity
345 contractile proteins	748 tumor virus expression	
	793 culture growth of fibroblasts	

developed in microbiology. In this case the link is the result of one field using the methods of another. The techniques of microbiology are also important to orthopedics in the diagnosis of bone infections, caused by organisms such as *Staphylococcus aureus*, and their treatment with antibiotics. Orthopedics links to the large central area of biomedical and physical science, and this is due to interest in various kinds of bone analysis which have found application in biomedical re-

search areas such as platelet deposition in bone marrow and rheumatoid arthritis.

The central area of biomedical and physical science has links to nearly all peripheral areas on the map. The 'double' links will be discussed here. Its link to pharmacology is in part due to basic research on the active transport of ions across cell membranes and the substances which control that process. Its link to chronobiology is the study of how circadian rhythms control the release of sub-

Figure 6. 1983 C4 cluster map, cluster no. 0002: social, behavioral, and biological sciences.



stances in the body such as the hormone prolactin. Its link to thermal biology concerns how blood flow regulates body temperature.

The link of the central biomedical and physical science cluster to physical topics begins with chromatography at the top of the map. This is due to a method in analytical chemistry, high performance liquid chromatography, used to study the concentration of drugs and toxic chemicals in the body. Another physically oriented link is to quantum optics on the left, which concerns basic and applied aspects of multiphoton interactions with matter, specifically as these relate to the development of high powered lasers undoubtedly relevant to recent 'star wars' research. Back in the central region of the map, the link from biomedical and physical sciences to social, behavioral, and biological science is the common concern with complex systems, as manifest in economic behavior, models of the brain, and the statistical mechanics of so-called 'self-organizing' systems. Complex systems appear to be an area of shared interest among social, physical, and biological researchers.

We will now switch over to the social, behavioral, and biological sciences cluster in the center and follow its links to other areas.

First we arrive at crisis medicine to the lower right, which is linked to social science by discussions of policy-related health statistics, including the incidence of diseases and accidents, use of medical technologies, and health care costs. Moving toward the left from social, behavioral, and biological science we arrive at agriculture via the ecology of plant populations, this link having a theoretical emphasis on the biological side and an applied emphasis on the agricultural side. Moving left from agriculture we arrive at earth science via the dynamics of soil, including erosion, decomposition of biomass, and sediments. Along this link the emphasis shifts from the short term perspective of agriculture to the longer term (geological) perspective of earth science. Finally, the path leads from earth science to materials science by a concern with wave motion in solids. The shifting interest along this link appears to be one of scale: earth science concerned with motions of solids on a large scale (e.g., tectonics), and materials science concerned with behavior of solids on a smaller scale.

This walk along a minimal spanning tree through the global map has shown that the links which bind the research areas have a specific, content-related character and rep-

resent substantive shared interests. It should be noted, however, that the nature of co-citation links between such macro-clusters is quite different from the links encountered at the document level, for example, on a CI map. At this low level it is possible to isolate single concepts and specify the nature of each document-to-document link with little ambiguity. At the macro-structural level, however, multiple reasons for association are the rule rather than the exception, and generalization is more difficult.

In the previous description we have seen that methods sometimes are the basis of the linkage, and sometimes empirical data. Occasionally one area provides theory, and the other area applies it to a practical problem. It is often difficult to ascertain the 'direction' of dependence or knowledge transfer, i.e., whether one area is the source, the other the recipient of information. In most cases it is possible to see a mutual benefit to the relation, an exchange of physical or intellectual resources. In some cases it is possible to discern a differential concern or interest across the link regarding the topic: e.g., basic *versus* applied, long range *versus* short range perspective, large scale *versus* small scale, social *versus* physical, methodology *versus* data, etc.

Some general findings of our mapping work are that research effort is not uniformly distributed across fields or countries, but rather obeys well known hyperbolic laws. In structure this is reflected in few rather large areas and many smaller ones, often arranged in center-periphery patterns. This is not a dramatic finding since we have known for some time that scientometric distributions are almost invariably hyperbolic. What we did not appreciate was how maps of science would have to reflect this mix of few large and many small areas in their structure. We also find that peripheral areas are often those with

applied or technological emphasis, while central ones are usually academic fields of basic research.

We have also found that by analyzing linkages between areas, it is possible to discern the nature of the relationships among them, in terms of scientific problems involved, the resources exchanged, and the different perspectives of the scientific groups involved. The next step in our mapping work will be to relate and coordinate maps derived from different time periods, so that changes in the micro- and macro-structures may be analyzed. Then we will be doing a *new* kind of history of science.

#### References

- [1] E. Garfield, I.H. Sher and R.J. Torpie. *The Use of Citation Data in Writing the History of Science*. ISI Monograph (Institute for Scientific Information, Philadelphia, 1964).
- [2] D.J. DeSolla Price, Networks of scientific papers. *Science* 149 (1965) 510-515.
- [3] H. Small, Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Amer. Soc. Information Sci.* 24 (1973) 265-269.
- [4] D.J. DeSolla Price, The Science of Scientists. *Medical Opinion and Review* 1 (10) (1966) 88-97.
- [5] S.C. Bradford, *Documentation* (Crosby Lockwood and Sons, London, 1948) 137.
- [6] V. Bush, Science—the endless frontier. A report to the President on a Program for Postwar Scientific Research, July 1945.
- [7] P. Gould and R. White, *Mental Maps* (Harmondsworth, England: Penguin Books, 1974).
- [8] D. Crane, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (University of Chicago Press, Chicago, 1972).
- [9] H. Small and E. Sweeney, Clustering the "Science Citation Index" using co-citations. I. A comparison of methods. *Scientometrics* 7 (1985) 391-409.
- [10] G. Salton and D. Bergmark, A citation study of computer science literature. *IEEE Trans. Professional Communication*. PC-22 (1979) 146-158.
- [11] H. Small, E. Sweeney and E. Greenlee, Clustering the "Science Citation Index" using co-citations. II. Mapping science. *Scientometrics* 8 (1985) 321-340.
- [12] J.B. Kruskal, Multidimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis. *Psychometrika* 29 (1964) 1-27.
- [13] H. Small, Citation context analysis. *Progress in Communication Sciences* 3 (1982) 287-310.