Back to Contents Page

Chapter Eight

Mapping the Structure of Science

Conceptually, there is not much distance between diagraming the intellectual structure of an historical line of research and doing the same thing for the whole of contemporary science. If, as the DNA-history study showed (1), citation analysis can be used to define the intellectual relationships between past research events, there is no obvious reason why it cannot be used to do the same thing for current ones. In other words, what the DNA-history study validated was not just a methodology for exploring the history of science but a methodology for defining the intellectual structure of science, past or present.

The structure of science is an intriguing subject to those who study science as a system. Beyond the intuitively comfortable hypothesis that science is a mosaic of small units, rather than a structural monolith, there are many more questions than answers. What is the nature of the basic units in the mosaic structure? How do they relate to each other? Are the intellectual and social structures similar? Are they made up of the same basic units? What is the relationship between them? Is there a variety of configurations at the infrastructure level? How dynamic are the configurations? Is there a relationship between configuration and research performance? Can structural analysis help us make more effective science policy decisions?

A number of studies during the first half of the 1970s began examining some of these questions in a systematic manner. Most of them provide evidence that specialties are the basic intellectual and social unit of the scientific mosaic. Crawford (2) described how the specialty of sleep research is built around the communications between a small group of key individuals and research centers. Griffith and Mullins (3) have shown that small, socially coherent groups are capable of producing major changes within disciplines. Crane's work on specialties in rural sociology and mathematics has led her to the idea of specialties forming a structure that can be mapped (4).

As might be expected, citation analysis has played a prominent and productive role in the attempt to define the structure of science. Building on the conceptual foundations of the literary model of science that was validated by the DNA history study, Price has used citation patterns to explore the structure of physics (5) and one of its specialties (6). Narin, working at a more general level, is using the citation patterns between journals to define the disciplinary structure of science (7). Goffman (8), Jahn (9), and Small (10) have all shown that a specialty is defined by a few critically important papers that appear early in its history. Small and Griffith (11) (12) have gone so far as to produce a map that showed all the high-activity specialties in the natural sciences. Small, on his own, has extended this work to produce a similar map of the social sciences (13) and a five-year series of maps that show the evolution of a single, biomedical specialty, collagen research (14).

The mapping studies by Small and Griffith deserve special attention. They represent the most sophisticated attempt made, as of the mid-1970s, to use citation analysis to define the structure of science on the scale and at the level of detail needed for science policy purposes. On the one hand, their method seems capable of depicting all the major scientific specialties in a single, coherent structure. On the other hand, it offers a range of resolution broad enough to permit the detailed examination of any substructure level that is appropriate to the questions an investigator may choose to pose. This combination of scale and variable resolution produces a functional capability, for analytical purposes, that is analogous to an automated design system in which a computer is used to display a product at all production levels, from final configuration down through the entire hierarchy of subassemblies. If the structures depicted by the Small-Griffith approach are validated as being reasonably realistic, the search for the structure of science will end, and the development of structural analysis techniques that produce functional insights will begin.

CO-CITATION CLUSTERING

The first study by Small and Griffith (11) was designed to test two hypotheses. One was that science is made up of a structure of specialties that can be defined by objective means. The other was that a particular citation measure of the common intellectual interest between two documents was a practical way of defining the structure.

The measure they used was co-citation strength, which is the number of documents that have cited a given pair of documents. Co-citation strength is a creative reversal of Kessler's bibliographic coupling concept (15), which uses the number of references a given pair of documents have in common to measure the similarity of their subject matter. The shortcoming of bibliographic coupling in structural studies of science is that the structure, presumably, is dynamic over time, whereas bibliographic coupling is a fixed measure. (Once a document is published, its references do not change.) In contrast, co-citation strength reflects the frequency of being cited, which is a characteristic that is variable over time. The rationale behind the use of co-citation strength was that its variability was caused by shifts in research focus and relationships (16). I.V. Marshakova also noted this characteristic in her paper on using reference citations for literature classification (17).

To test these hypotheses, Small and Griffith worked out the method shown in Figure 8.1 for identifying clusters of papers that are linked by specific levels of co-citation strength. The processing is done by computer. The initial input to the processing cycle is a calendar increment (quarter, semiannual, annual) of the *SCI* or *SSCI* data base, depending on whether the study is concerned with the natural or social sciences.

Processing begins with extracting a highly cited subset of the data base, which includes both reference and source citations. Setting the threshold of citation frequency that will qualify documents for inclusion in the subset is an important matter of judgment. On the one hand, the threshold should be set low enough to pick up all the documents that could be considered, by the measure of citation frequency, the core of the scientific literature. This core material is representative of the full spectrum of significant research activity. On the other hand, the threshold must be set high enough to exclude documents that do not add anything to the definition of research areas, and may even obscure some of them.

Regardless of the citation-frequency threshold set, the product of the first processing step is a citation index of highly cited documents and the papers that cited them. In the second step, this material is sorted to produce a source-item index of the highly cited documents. In other words, the highly cited documents are organized by the source papers that cited them, so co-citation links can be identified.

All pairs of documents that have been cited by the same source paper (co-cited) are then extracted, and the documents in each pair are arranged in alphabetical sequence. The result is a list of document pairs in which each pair appears in only one of the two sequences that are possible: as AB, but not as BA. The pair list is then put into alphabetical order, which brings together all identical pairs.

To reduce the number of records that must be processed, the alphabetically ordered pair list is consolidated. The number of times each pair appears is counted, the total is recorded next to the first record of the pair, and all duplicate records are deleted.

The next step is to duplicate the consolidated pair list, but with the document sequence of each pair reversed: AB becomes BA.

The two consolidated pair lists are then combined and put into alphabetical order, producing a master list of co-cited pairs on which all pairs that have a single document in common are batched together.

The last step in the procedure is running the master list of co-cited pairs through a clustering routine that aggregates clusters of documents by sequentially linking together all pairs that have at least one document in common. Starting with one pair, say AB, the clustering routine links it to all pairs that include A. It then links to that aggregation all pairs that include a given document that has been linked to A. That procedure is repeated, document by document, until there are no pairs left on the master list that have one of the aggregated documents in common.

At that point, the routine has produced a cluster of documents that are related

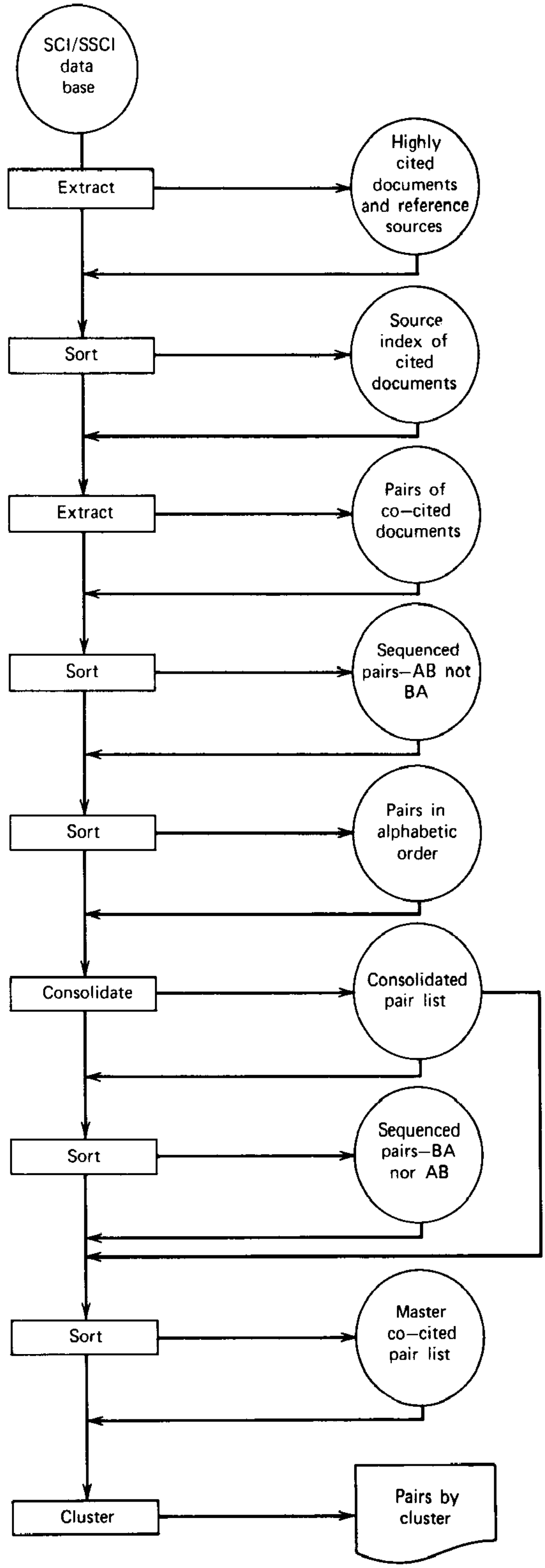

Figure 8.1 Functional diagram of co-citation clustering method.

either directly or indirectly, by co-citation. It then goes through the same operation for each of the unassigned pairs until all the entries on the master list either have been clustered or have been found to be isolated from all the clusters.

The output of the clustering routine is a printout of the pairs contained in each cluster, along with their co-citation strengths, and a list of the isolated pairs that do not fit into any clusters.

The resolution of the clustering operation, in terms of the research scope defined by the papers in individual clusters, can be controlled by setting a threshold for the co-citation frequency that qualifies a pair for inclusion in the clustering routine. Every pair on the master list has a co-citation frequency of at least one, so a co-citation threshold of one would qualify every pair on the list for clustering. As the threshold is raised, however, fewer pairs qualify, and membership in a cluster becomes more exclusive. Because a reduction in the number of pairs in a cluster reduces the potential for subject diversity, raising the cluster threshold has the effect of defining the intellectual scope of the cluster more narrowly—or, increasing cluster resolution.

NATURAL SCIENCE STRUCTURES

The first application of this methodology was in the natural sciences. The data base used was the first quarter of the 1972 *SCI,* consisting of 867,600 reference citations and 93,800 source citations. The source papers came from approximately 2400 journals, which cover just about all the disciplines in the natural sciences.

The threshold set for selecting the subset of highly cited documents from which the master pair list would be derived was a citation frequency of 10. The subset obtained with this threshold contained 1832 reference documents and 16,927 source papers. Out of a possible 1.7 million pairs that could have been formed from the reference documents, they yielded only 20,414. The yield as a percent of potential, which is a measure of the integration of the subset, worked out to a low 1.2%, indicating that the intellectual relationships in the subset were loosely structured. The co-citation strength of the 20,414 pairs ranged from 1 to 81, with 1.78 being the average—another indication of the looseness of the subset structure. On the other hand, all but 18 of the reference documents were included in at least one pair—an indication that, though the structure may be a loose one, there is a definite structural potential.

The subset was clustered at three co-citation threshold levels: 3, 6, and 10. Figure 8.2 shows the number of pairs that qualified for clustering at each of the levels, the number of unique documents in the pairs, and the number of unique documents that did not qualify (isolated cases). Figure 8.3 shows how the clusters at each level were distributed by size and the number of cited documents in each size category. Clusters consisting of only two documents are, of course, not clusters at all, but individual pairs that met the co-citation threshold but did not link to any other pairs. Ignoring them, the number of clusters formed at each level was:

Level	*No. of Distinct Pairs*	*No. of Documents Paired*	*No. of Isolated Documents*
3	3,067	1,310	522
6	791	594	1,238
10	213	193	1,639

Figure 8.2 Co - citation pairs that qualified for clustering.

Cluster Size in Papers	NO. OF CLUSTERS			NO. OF CITED ITEMS		
	Levels			*Levels*		
	3	*6*	*10*	*3*	*6*	*10*
2	71	60	21	142	120	42
3	13	17	3	39	51	9
4	9	11	6	36	44	24
5	7	5	3	35	25	15
6	5	2	4	30	12	24
7	1	1	1	7	7	7
8		3			24	
9	2	1		18	9	
10	1	2		10	20	
11						
12	1	1		12	12	
13						
14						
15	1			15		
16						
17		1			17	
18						
19						
20						
26		1			26	
27		1			27	
32	1			32		
41	1			41		
72			1			72
92	1			92		
200		1			200	
801	1			801		
Total	115	107	39	1,310	594	193

Figure 8.3 Distribution of clusters by number of cited documents.

- Level 3: 44 clusters.
- Level 6: 47 clusters.
- Level 10: 18 clusters.

The largest cluster at each of the three levels consists of documents on biomedical research: 801 at level 3, 200 at level 6, and 72 at level 10. This cluster accounts for 61% of all the documents that qualified for clustering at level 3, 34% at level 6, and 37% at level 10. The large drop that takes place in the percent of documents accounted for by the biomedical cluster when the threshold is raised from 3 to 6 suggests that the cluster is rather loosely linked. The fact that the percentage does not drop further when the threshold is raised to 10, however, suggests that a number of the links are quite strong.

Figure 8.4 identifies the 10 largest clusters at level 3. Each is described by a name derived from the titles of its documents, the number of documents and pairs it contains, its integration level (percent of potential pairs), and the average publication date of its documents. The three largest clusters (at all levels, in fact) were *biomedicine, chemistry,* and *nuclear physics.* Again, low integration levels imply that they are very loosely structured. This time, the inference was confirmed by an examination of the documents: each cluster was found to contain several sets of documents on subjects considerably more specialized than the cluster designation. The list also shows clearly the relationship between cluster size and specialization: as the cluster size decreases, the subject matter of its documents becomes more specialized.

With the exception of *biomedicine* and *chemistry,* which describe disciplines, the names assigned to the clusters formed at level 3 all identified specialty areas of research. Since the names were derived from the titles of cluster documents, this finding suggests that the clusters represent the specialties that some researchers think are the basic structural units of science. Analysis of the documents in clusters formed at the other two levels produced similar results.

The relationship between clusters and specialties was further examined by analyz-

	No. of Documents	*No. of Pairs*	*Percentage Connectedness*	*Average Date*
Biomedicine	801	2,205	0.69	1963.3
Chemistry	92	291	7	1962.5
Nuclear physics	41	59	7	1964.3
Particle physics	32	99	20	1968.1
Australia antigen	15	70	66	1969.0
Crystal structure of enzymes	12	30	45	1968.0
Plate tectonics	10	35	78	1967.8
Virally transformed cells	9	25	69	1964.2
Nuclear magnetic resonance	9	8	22	1957.0
Neurophysiology of vision	7	14	66	1963.1

Figure 8.4 Ten largest clusters at level 3.

ing the structure and interactions of a few of the clusters. For this analysis, the network of documents in the *nuclear physics* and *particle physics* clusters formed at level 3 were drawn as simple block diagrams. This had the effect of dramatizing the structural difference that is indicated by the disparate integration levels of the two clusters (Figure 8.4). *Nuclear physics,* which had an integration level of only 7% was quite easy to diagram (Figure 8.5), using blocks to represent its documents and lines between the blocks to represent the strength of the co-citation linkages between documents. However, when it came to the *particle physics* cluster, with an integration level of 20%, this diagraming method was only partially successful (Figure 8.6). Twenty-one of its documents were so tightly linked that it would have been impossible to distinguish between the connecting lines.

To find out if the structures of the two clusters were logically consistent, word profiles were constructed for each document from the titles of the papers that cited it. The rationale behind this method of analyzing the contents of the cluster documents is the same one that underlies the use of citation counts as a measure of research utility: the usefulness of a document is an attribute conferred by the authors who cite it. It follows, then, that an effective way of generally identifying the useful content of the document is by defining the subject matter of the citing papers. Analyzing the title words of the citing papers is the easiest way of doing that.

The profile for each cluster document consisted of the four words most frequently used in the titles of the papers that cited it. The frequency rate of each word also was included in the profiles. With these profiles in the appropriate diagram boxes (Figures 8.7 and 8.8), it was possible to analyze the logical consistency of the clusters. What was being looked at were word patterns: consistency between linked boxes, differences between unlinked ones, and a logical progression of change over a series of linked boxes.

The patterns found seemed to be logically consistent for the specialties of nuclear and particle physics. For example, in the *nuclear physics* cluster, the groups of papers linked together at the top, slightly to the right of center, all have profiles in which the word "scattering" is prominent. As you move down the vertical line of linked documents in the center of the cluster, the profiles describe a progression that goes from nuclear reaction to fission. Conversely, the profiles for unlinked documents are properly dissimilar.

The same type of consistency exists in the *particle physics* cluster. At the bottom, the profiles describe work on current algebras and broken chiral symmetry. As you move up, they introduce applications of the Veneziano model; and, at the top, in the large box containing the tightly linked documents that defied diagraming, they shift to the vocabulary used to describe work in inclusive reactions and the internal structure of elementary particles.

More evidence of the physics clusters being representative of the specialties implied by their names was provided when they were both reclustered at a co-citation threshold level of 1. At that level of commonality, the two clusters were linked at the points shown by the dashed lines on Figures 8.5 through 8.8. Most of the connections were made between the "scattering" work in *nuclear physics* and the loosely structured tail of *particle physics.* The appropriateness of this connection was con-

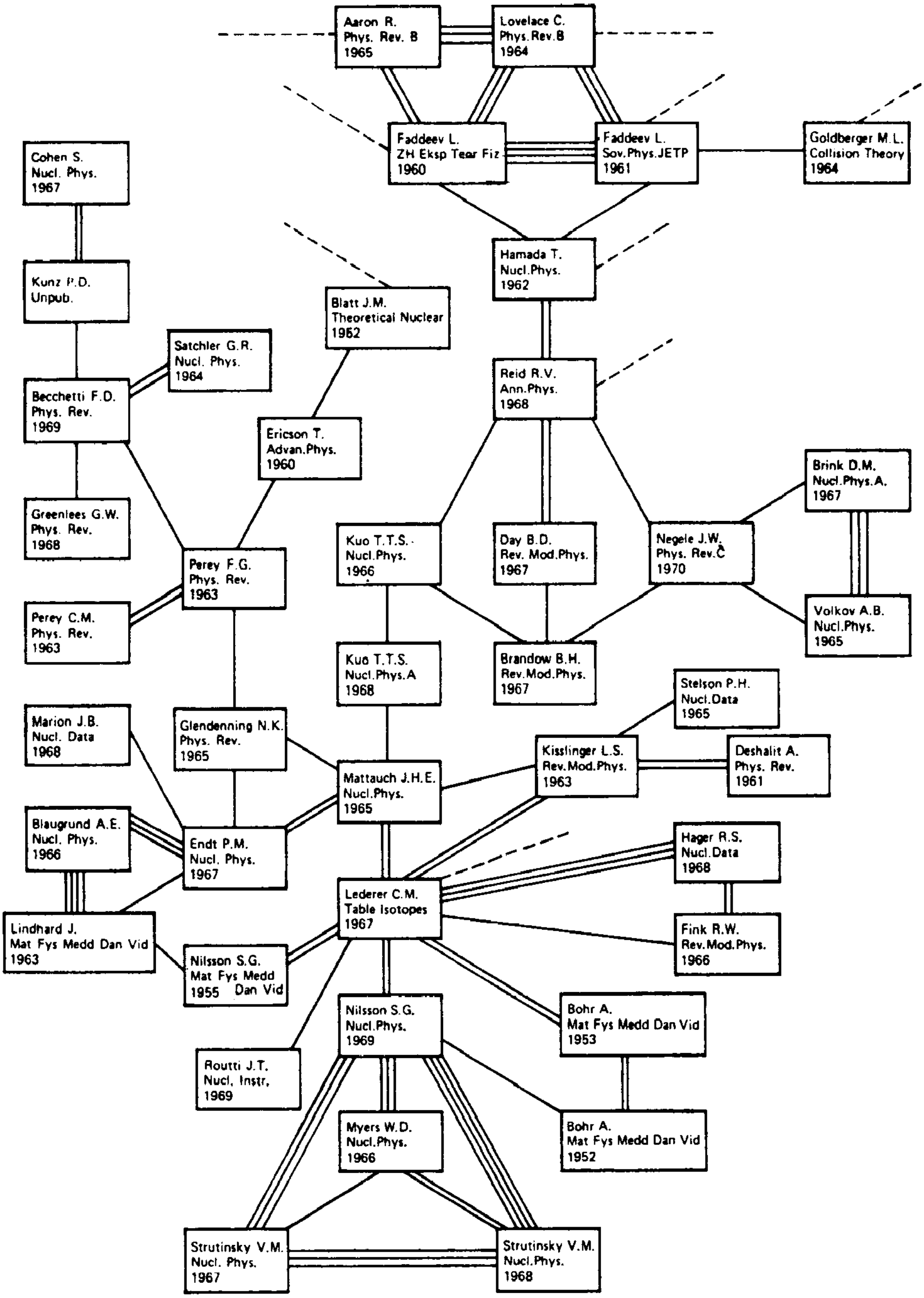

Figure 8.5 Block diagram of level 3 *nuclear physics* cluster. Each box represents a cited document, which is identified by the name of the first author, the title of the journal or book, and the year. Connecting lines between boxes indicate co-citation links, with the number of lines proportional to the strength of the linkages according to the following scale:

Number of Lines	*Co-citation Strength*
1	3
2	4–5
3	6–8
4	9–12

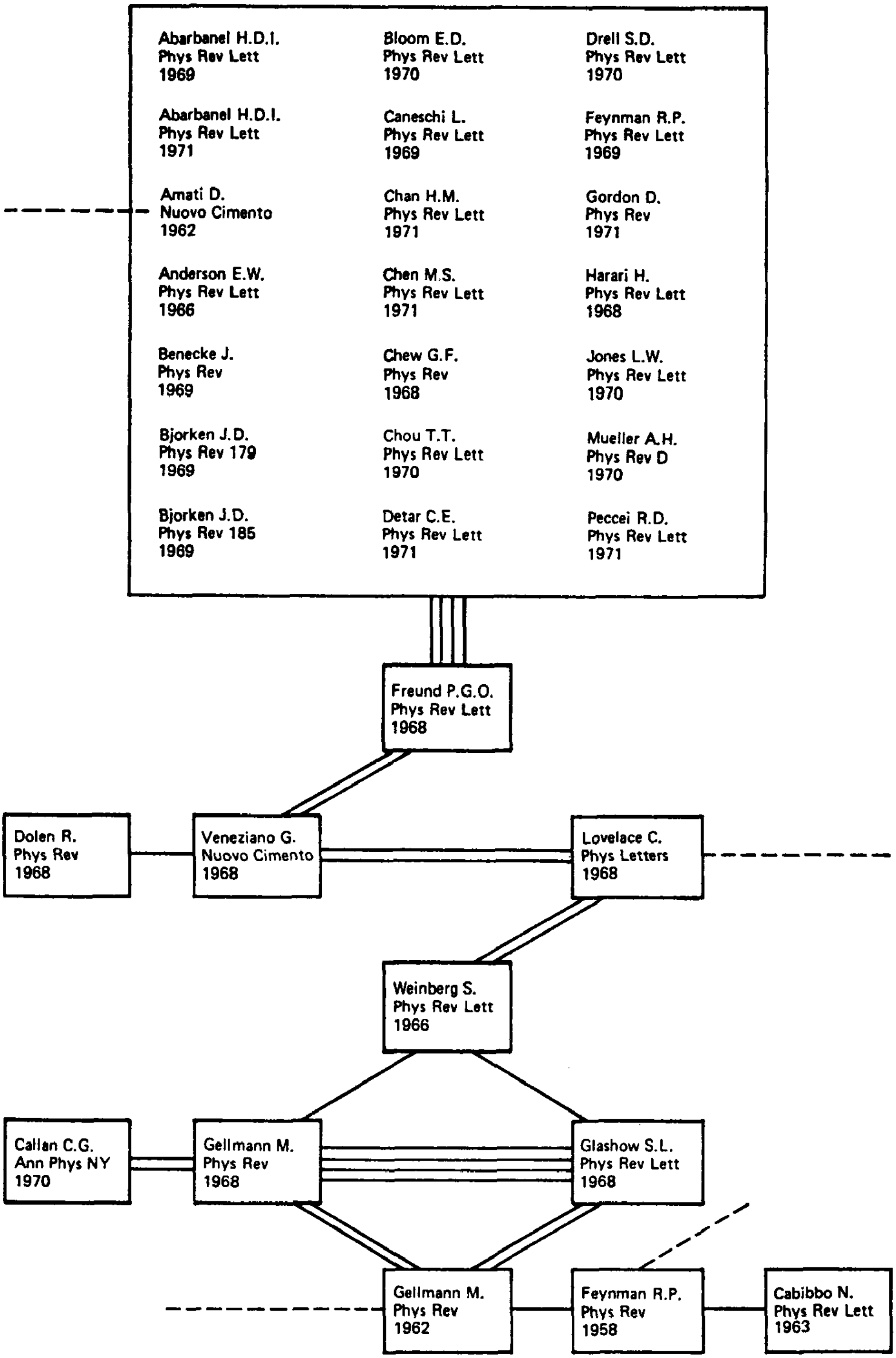

Figure 8.6 Block diagram of level 3 *particle physics* cluster. Documents and co-citation linkages are indicated as in Figure 8.5, with the exception of large box at the top. The co-citation linkages between the 21 documents listed in that box were too strong to depict with lines. The dashed lines indicate co-citation linkages to the *numclear physics* cluster below level 3.

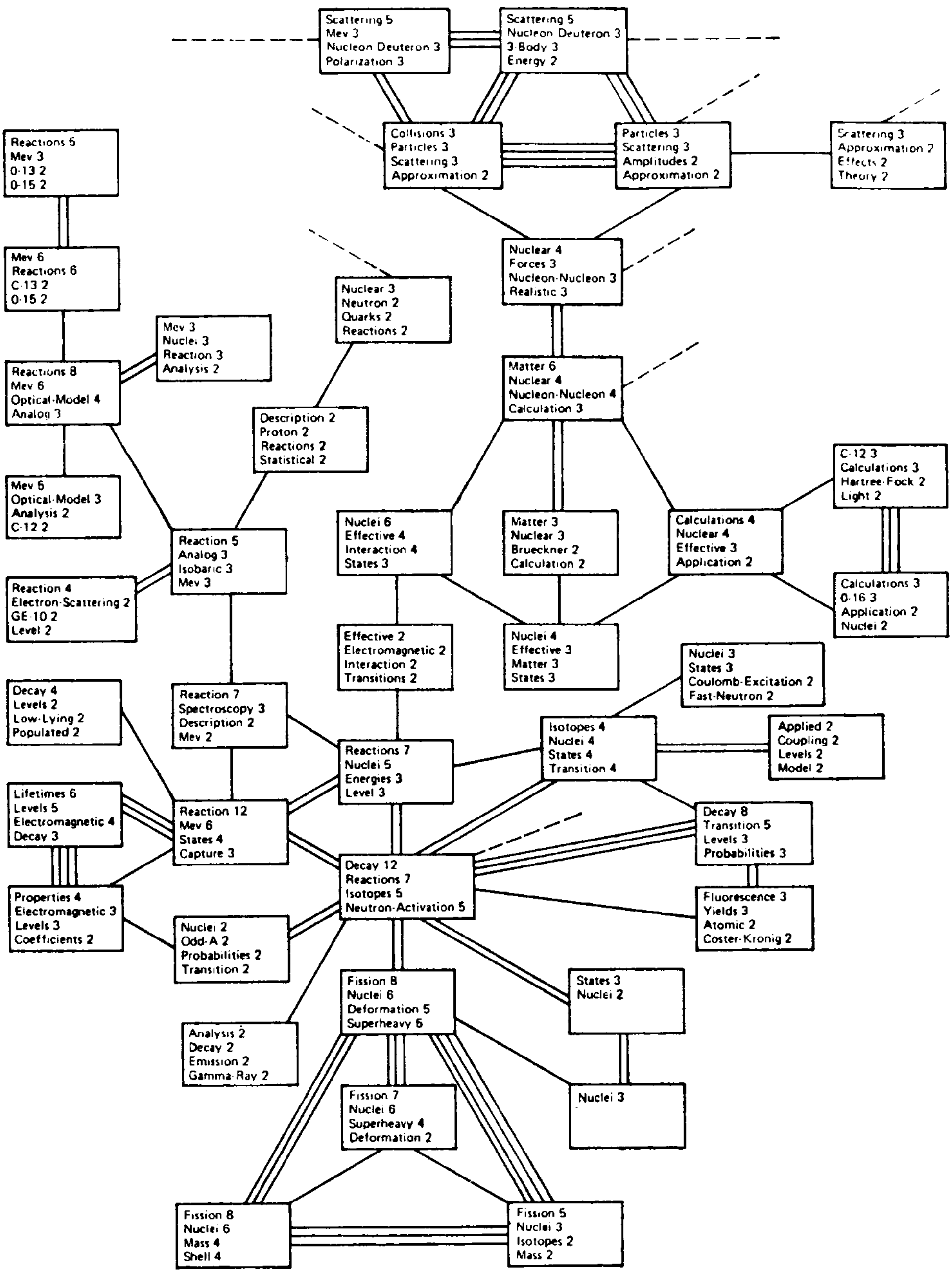

Figure 8.7 Word profiles of level 3 *nuclear physics* cluster. Numbers following the words indicate frequency of occurrence.

firmed by the author of one of the articles that formed the connection—a review article on final-state interactions—who wrote, "The main subject of this review article is a chapter of physics which is recognized as common ground for nuclear physics and what is currently known as particle physics."

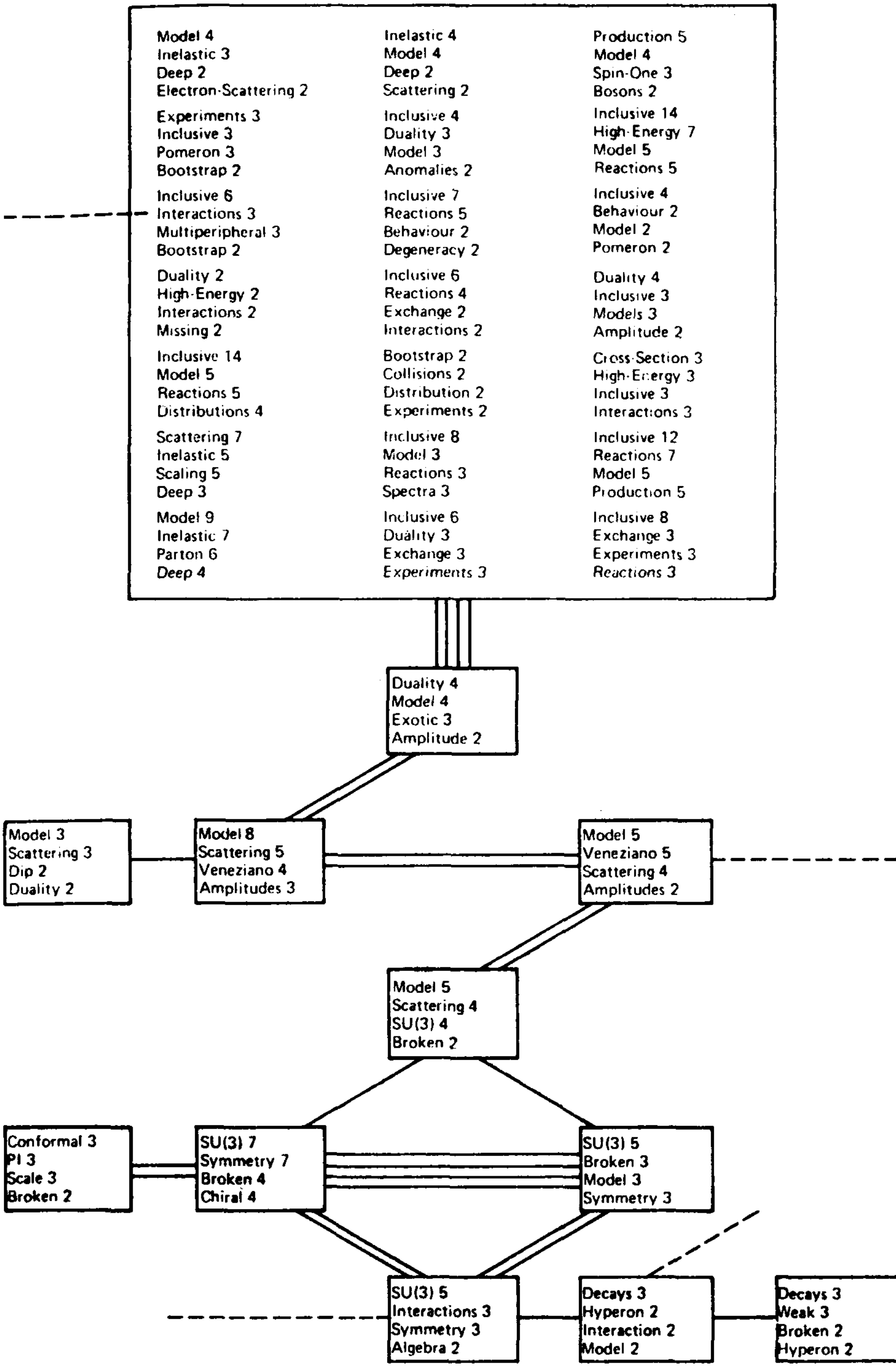

Figure 8.8 Word profiles of level 3 *particle physics* cluster. Numbers following the words indicate frequency of occurrence.

A similar attempt to see where one of the level 6 clusters connected at a lower threshold level involved a cluster named *reverse transcription*. That, too, tied into an appropriate structural point: earlier work in molecular biology.

The *biomedicine* cluster, being the largest of the two anomalies that were disciplinary in scope, also was examined in detail. Though it was steadily reduced in size as the co-citation threshold was raised, it seemed to resist attempts to break it into clusters as compact and sharply focused as most of the others. At level 10, it still contained 72 papers that had more of the diversity of a discipline than the focus of a specialty. An analysis of the highly co-cited papers at that level revealed the problem: all but two of the 31 documents that had the highest co-citation rates were methodology papers.

Several strategies, involving the removal of increasingly large subsets of the documents from the clustering data base, were tried to determine the structural role of the methodology documents. Only when all of them were removed and the remaining documents were clustered at level 4 did the individual biomedical specialties emerge. The result was 74 clusters, each one with the intellectual focus of a specialty.

The analysis of the biomedical cluster showed, then, that it had two structures: a conceptual one of specialties and a methodological one of techniques, with the methodological one functioning as a superstructure that tended to link together specialties that had little more in common than a set of shared techniques.

The finding of the dual structure in *biomedicine* was, in comparison, the least of what the study produced. Of more fundamental importance were the findings that co-citation clustering succeeded in identifying groups of papers that were linked by a common intellectual interest, and that the clusters of documents produced by that method seemed to correspond to the specialties that are said to make up the mosaic structure of science. In fact, the study provided the first large-scale evidence of the notion that science is a network of specialties, and that the network can be viewed through the literature.

The study's evidence in support of these two assertions is quite strong. The clustering patterns found certainly were not inevitable. There was a distinct possibility that none of the highly cited papers would be co-cited, or that none of the co-cited pairs would link together. Alternatively, the study could have found the interaction between all areas of research (and the co-citation links between documents) to be so strong that all the highly cited documents formed one gigantic cluster even at high co-citation threshold levels.

The clustering method used in the study, in fact, was biased in favor of producing the second alternative. Because it clusters on the strength of only one link between two pairs, it has a tendency to string together objects that have little in common. Yet, this tendency was evident only at the very low co-citation levels; through most of the threshold range, there was a realistic amount of discrimination. There is no doubt, therefore, that the clusters produced were intrinsic to the data, rather than the methodology.

Though the evidence for the clusters being representative of specialties was considerably less conclusive, it certainly was consistent and encouraging. The clusters of documents fit expectations of how specialties should look and behave. Typically, they consisted of a core of discovery papers, surrounded by others describing work built on the original discoveries. The word analysis showed that their internal structure was logical. And the way the intellectual scope of clusters varied with the co-citation threshold level was consistent with the theory that the specialties of science create a heirarchical structure that forms an all-encompassing mosaic pattern of interaction at the lowest level.

Aside from the obvious importance of these structural findings, the study was significant also for the methodology it demonstrated. Citation analysis has two distinct advantages over the more traditional methods of exploring the structure of science through its literature. One is that much of it can be automated—a characteristic that makes it practical to study science on a scale large enough to distinguish between special and universal characteristics and to make valid comparisons from one specialty to another. The functional practicality of the method also makes it feasible to repeat studies frequently enough to closely monitor changes over time.

The second advantage of citation analysis is its objectivity. Other methods for studying specialty behavior from the literature are based on subjective judgments about what comprises the literature of the specialty. In co-citation clustering, this judgment is made algorithmically from the hard data of citation relationships, which are specified on a massive and inclusive scale by the population of publishing scientists.

MAPPING SPECIALTIES

The success in using co-citation clustering to identify specialties led to the next logical step of using the same technique to map them (12). To do this, Small and Griffith extended their methodology in the way shown in Figure 8.9.

Mapping the specialties picks up where defining them leaves off. The clustering routine can produce a list of clustered documents organized by cluster number. This list is matched against the master list of co-cited pairs, and all the documents on the pair list that appear on the cluster list are tagged with the appropriate cluster number. All the rest, which are part of pairs that have not been co-cited frequently enough to be clustered, are left untagged. The result of the matching and tagging operation, therefore, is to divide the pair list into three classes: those pairs in which both documents have the same cluster number; those in which the two documents have different cluster numbers; and those in which only one of the documents, or neither, has a cluster number.

Document pairs with the same two cluster numbers are those co-cited frequently enough to have an intracluster relationship. Those that have two different cluster numbers have not been co-cited frequently enough to qualify for an intracluster relationship, but they do have an intercluster relationship. Those pairs in which only one document, or neither, has a cluster number are those that have no relationship at all with the clusters formed at the co-citation threshold used.

The first two classes, of course, are the ones used in mapping the specialties.

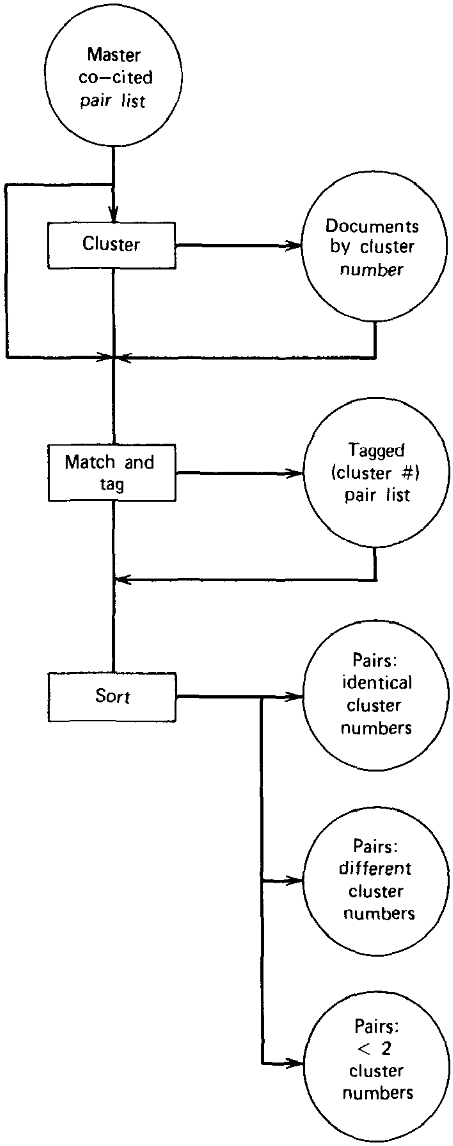

Figure 8.9 Methodology of mapping specialties.

Those pairs with intracluster links represent the contents of the individual clusters. Those with intercluster links provide a co-citation measure of the relationship between clusters. Together, they can be used to map the specialties in a number of ways.

Figure 8.10 is one of the ways. A simple network diagram, its nodes represent individual clusters. The number in each node corresponds to the number of the cluster. Those clusters that contain three or more documents are identified in Figure 8.11, which also shows the number of documents they contain. The lines connecting the nodes represent co-citation links between the clusters. The figures beside the lines represent the strength of those links. Each of the figures is the cumulative

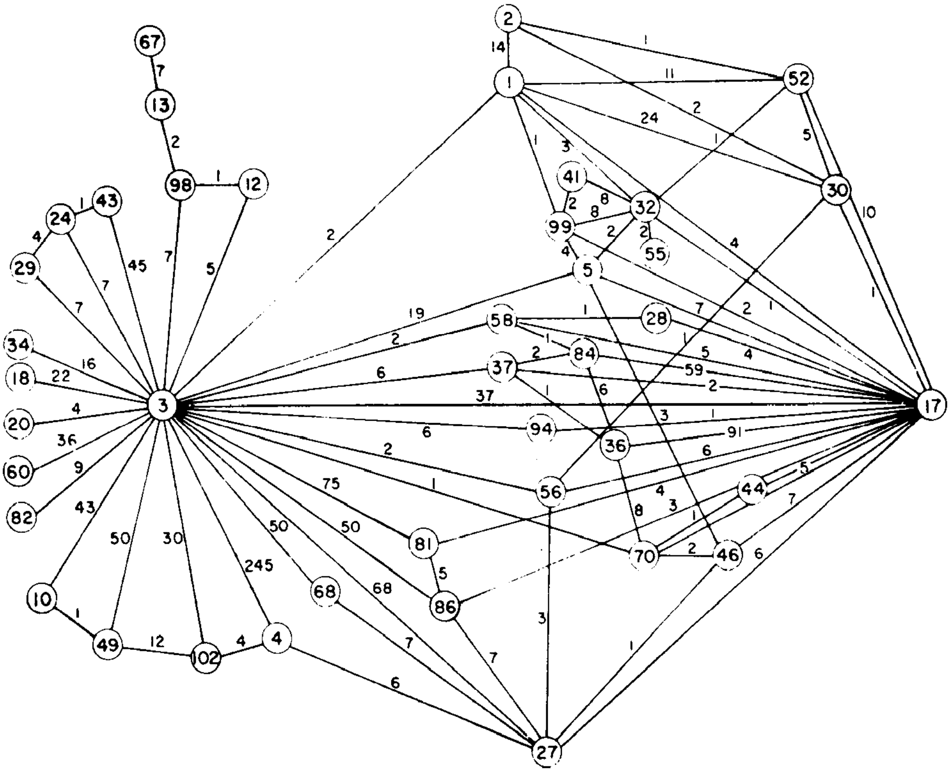

Figure 8.10 Network diagram of level 3 clusters. Circles represent clusters, which are numbered. The lines connecting clusters indicate co-citation links. The numbers on the lines indicate the strength of the links.

number of times unique pairs with a document in each of the two clusters have been co-cited. The co-citation strength of each unique pair that contributes to the total is, of course, lower than the co-citation threshold used for clustering.

The data base from which Figure 8.10 was produced is the same one used in the initial study to test the co-citation clustering technique: the first quarter of the 1972 *SCI.* The nodes are the clusters produced when all the documents in this data base that had been cited at least 10 times were clustered at a co-citation threshold level of 3.

The picture Figure 8.10 provides of the natural sciences is a logically comfortable one. It shows three disciplinary poles: *physics* (clusters #1 and #2, with 73 documents), *chemistry* (#17, with 92 documents), and *biomedicine* (#3, with 801 documents). Most of the specialties are connected to the *biomedicine* and *chemistry* poles and are small, which may be an intrinsic characteristic of scientific specialties. A large number of the ones that are connected to *biomedicine* have no connections to any other part of the natural sciences. In contrast, the specialties linked to *chemistry* are also connected to either *physics* or *biomedicine.* This pattern suggests that chemistry functions as a critical point of integration for much of the natural sciences—a hypothesis that is supported further by the very weak connection between *physics* and *biomedicine.*

Cluster No.	No. of Documents	Description
1	41	nuclear structure physics
2	32	particle physics
3	801	biomedicine
4	9	virally transformed cells
5	9	nuclear magnetic resonance (spin relaxation)
10	15	Australia antigen (viral hepatitis)
12	4	multivariate statistics
13	10	plate tectonics
17	92	chemistry (crystallography, molecular orbital theory, nuclear spin polarization)
18	3	bacterial susceptibility to antibiotics
20	3	kidney transplantation
24	3	psychology (arousal expectancy)
27	12	crystal structure of enzymes
28	4	radiationless transitions
29	7	neurophysiology of vision
30	4	nuclear reaction theory
32	5	solid state physics (theory of alloys)
34	5	renin
36	4	ligand field theory (charge transfer)
37	5	organic chemistry (substituent effects)
41	5	solid state physics (semiconductors, band theory)
43	6	Parkinsonism (treatment with L-dopa)
44	3	carbanion chemistry
46	6	nuclear magnetic resonance (paramagnetic shift reagents)
49	6	viral leukaemia
52	6	atomic physics (self-consistent fields)
55	3	solid state physics (critical phenomena)
56	5	computer science (minimization methods)
58	3	chemistry of singlet oxygen (photo-oxygenation reactions)
60	3	radioimmunology (fsh)
67	3	geology (origin of basalt magmas)
68	3	protein conformation (circular dichroism)
70	3	chemistry (dipole moments, Ising model)
81	5	transfer RNA
82	4	marihuana chemistry
84	4	chemistry (cycloadditions)
86	4	stereochemistry of nucleic acids
94	3	light scattering (biological cells)
98	5	numerical taxonomy (multidimensional scaling)
99	3	solid state physics (lattice theory)
102	4	cancer (detection with immunofluorescence)

Figure 8.11 Description of level 3 clusters with at least three documents.

There are only three obvious anomalies on this map of the natural science specialties, and all of them can be explained. *Plate tectonics* (#13) is attached, though indirectly, to *biomedicine.* This unlikely link between the earth sciences and biomedicine is created by a methodology [*numerical taxonomy, multidimensional scaling* (#98)] that is common to both. *Computer sciences* (#56) is linked, though weakly, to *biomedicine* (#3), *chemistry* (#17), *nuclear reaction theory* (#30), and the *crystal structure of enzymes* (#27), but not, as one would expect, to mathematics. The reason for this is that the citation threshold of 10 was too high to pick up very much of the mathematics literature, which uses references more sparingly than most of the rest of the natural sciences. That situation precluded the possibility of *computer sciences* connecting to anything other than its major application areas.

The third anomaly is the continued appearance of macroclusters that encompass several specialties. The *biomedicine* and *chemistry* clusters are this type. In their first study, when Small and Griffith investigated this phenomenon by analyzing the *biomedicine* cluster, they found that a superstructure of methods documents held together multiple substructures of conceptual papers that otherwise would have formed separate specialty clusters. In their second study, they extended their investigation by mapping the methodological superstructure and specialty substructures within the *biomedicine* cluster.

The methodological superstructure is shown in Figure 8.12 as a spatial display in which the documents are depicted as points that are plotted on a two-dimensional scale. The distance between points is inversely proportional to the strength of the co-citation links between them (the stronger the links, the shorter the distance). Pictured on the display are 28 of the 29 methods papers that were found to be holding together relatively disparate specialties. The missing paper is one by O. H. Lowry on a method for measuring proteins. It was left off the map because it was so highly co-cited with each of the other documents that all the points ended up in the same location when plotted on a two-dimensional, proximity scale.

The logical structure of the diagram can be derived from the titles of the documents involved. The six papers at the top, reading down from Karnovsky to Reynolds, all deal with electron microscopy. The three papers directly below them, by Bartlett, Chen, and Folch, are all on the measurement of phosphorous, particularly lipids. The measurement of lipids, as well as sugars, ACTH, and proteins, is the subject of the three papers in the center by Fiske, Gornall, and Dubois. The two papers to the right of center, lying above and below the horizontal axis, by Burton and Marmur, are on the measurement of DNA. The three papers by Ellman, Ornstein, and Davis, which lie on an arc beneath and to the left of the central group, are on the separation of organic materials by electrophoresis. Separation by molecular weight, using ultracentrifugation, is the subject of the large group of papers that straddle the vertical axis at the bottom. To the right of them are a number of older, classical methods papers.

There is, then, a definite logic to the structure. As shown in the lower right-hand corner of Figure 8.12, the subject matter begins at the top with electron microscopy methods for dealing with relatively large biochemical structures and progresses to methods concerned with the separation of very small structures (molecular in scale) at the bottom.

The major specialty substructures that were held together by the methodological superstructure are shown in Figure 8.13. When the 29 methods papers were removed from the data base, and the remaining 772 documents were clustered at a co-citation threshold level of 4, 74 clusters were produced. Forty-six of them, the ones shown in Figure 8.13, were interlinked at a co-citation strength of at least three. The size and

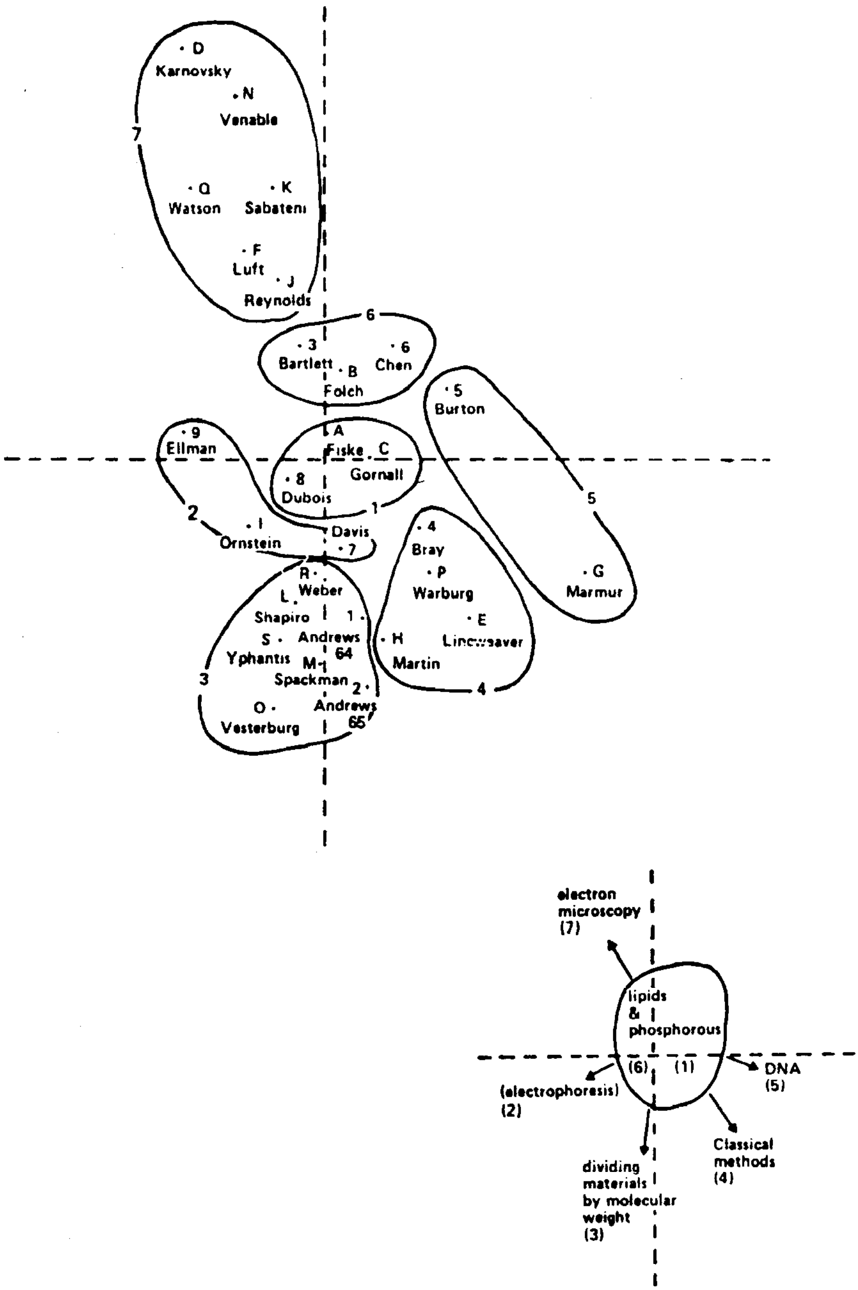

Figure 8.12 Spatial display of *biomedicine* methods papers. The papers are indicated by dots and identified by the name of the first author. The free-form shapes enclosing groups of papers represent clusters. The diagram at lower right shows the logical relationships between the clusters.

names of the interlinked clusters that defined at least four documents are shown in Figure 8.14. The names were derived from the titles of both the documents in the clusters and the papers that cited them.

The majority of the specialty clusters formed a very loose, treelike structure that, again, was logically consistent. For example, *immunology* (#9) was properly linked to both *cancer research, reverse transcription* (#10) and *structure of im-*

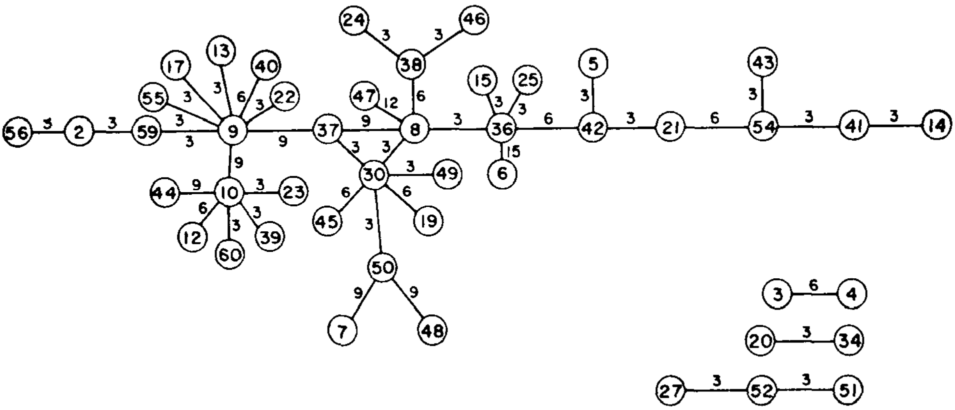

Figure 8.13 Network diagram of the major *biomedicine* specialty clusters. Circles represent clusters; the lines connecting circles represent co-citation links; numbers on the lines represent the strength of the links.

Cluster No.	*No. of Documents*	*Description*
1	15	fat transport in lipoproteins
2	9	radioimmunoassay of hormones
6	6	transport across cell membranes
7	4	fibrinogen, defibrination
8	34	protein structure, allosteric systems
9	65	immunology
10	57	cancer research, reverse transcription
11	11	human chromosomes, fluorescence patterns
12	4	mitochondrial DNA
13	6	lymphocyte stimulation
14	6	bactericidal capacity of leukocytes
15	8	muscle
16	5	digitalis intoxication
19	7	cellular hypersensitivity
20	16	platelet function
23	5	RNA polymerase, cyclic re-use
24	18	cyclic AMP
25	5	intracellular transport of secretory proteins
26	7	myosin structure
29	4	biomembrane lipids
30	15	secretory IGA, polypeptide chains
33	10	microsomal enzyme induction
36	17	erythrocyte membranes
37	14	structure of immunoglobulins
39	4	RNA metabolism
40	4	genetic control of antibody response
41	4	cell junctions, ultrastructure
42	4	water transport in biological membranes
44	4	bacterial DNA
47	6	carbohydrate metabolizing enzymes
51	4	glucogenic amino acids, metabolism
61	9	bio-statistics
62	17	biogenic monamines, effect on brain

Figure 8.14 Description of *biomedicine* specialty clusters with at least four documents.

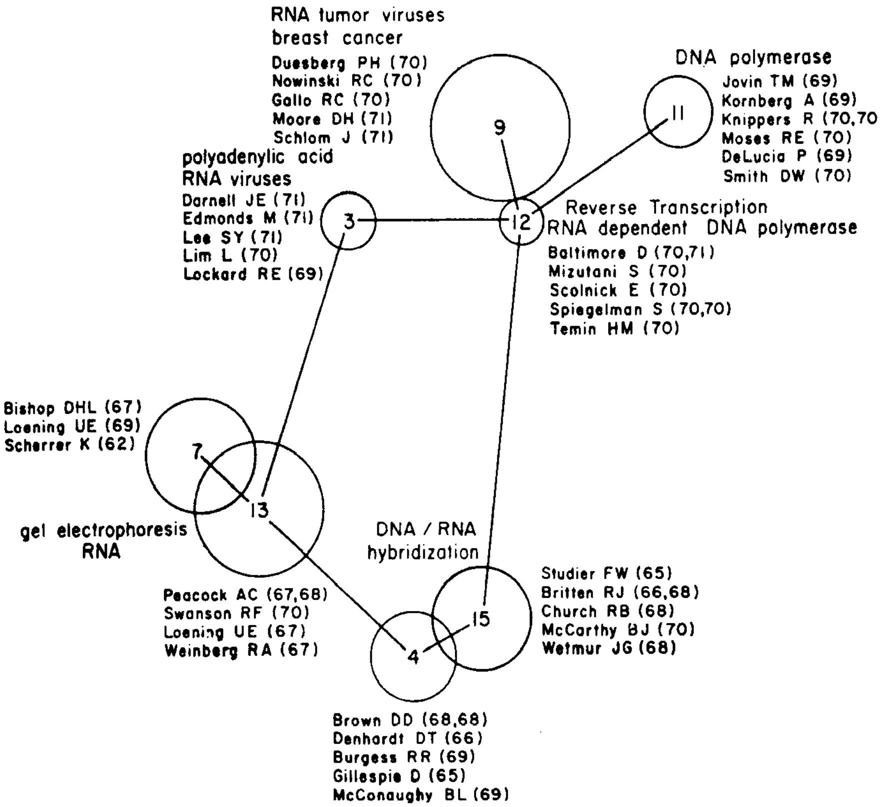

Figure 8.15 Spatial diagram of *cancer research, reverse transcription* macrocluster. Each circle represents a cluster. Cluster documents are identified by the name of the first author and the year of publication. Terms adjacent to clusters approximate the subject matter of the documents.

evidence that the structure of the biomedical clusters, and their behavior from one year to the next, matches the general perception of what is happening in the biomedical specialties. Since they were annuals, the data bases used for this study were considerably larger and more comprehensive than the quarterly one used for the initial clustering studies. The 1972 annual data base consisted of 2.6 million reference citations; the 1973, 2.7 million. A citation-frequency threshold of 15 was used to select a subset of documents that were paired and, then, clustered at a co-citation threshold level of 11. A total of some 900 clusters were produced from the 1972 and 1973 data, of which approximately 65% were identified as biomedical.

Figures 8.16 and 8.17 show, in map form, the biomedical clusters in each of the two years that contained at least three documents and were linked at a co-citation strength of at least 100. The parenthetical numbers in the boxes on the maps indicate the size of the cluster in documents, while the ones in the lines linking the clusters show the co- citation strength of the links.

The 1972 map (Figure 8.16) showed biomedicine as being dominated (in terms of number of documents) by four specialty areas: *RNA viruses, reverse transcriptase*

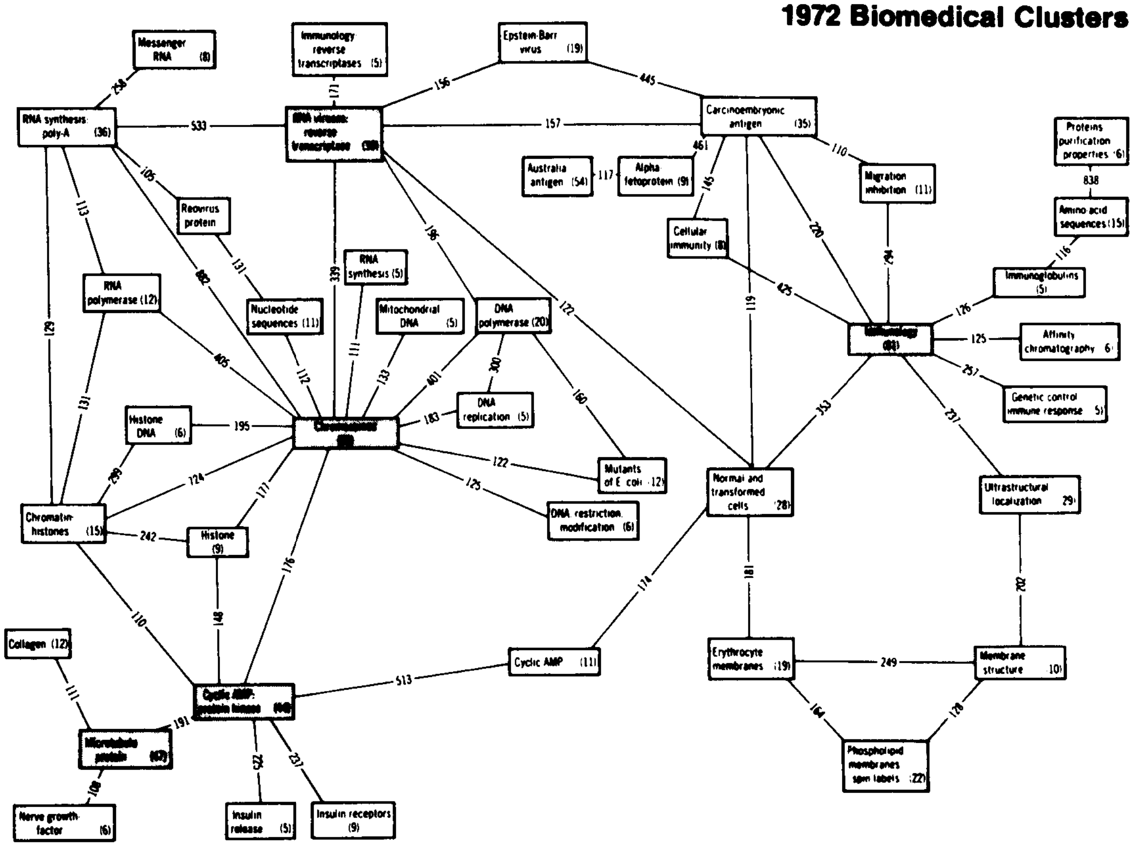

Figure 8.16 Map of 1972 biomedical clusters. Boxes represent clusters. Parenthetical numbers in boxes indicate the number of documents in clusters. Co-citation links between clusters are shown by connecting lines; strength of links is shown by numbers in lines.

1973 Biomedical Clusters

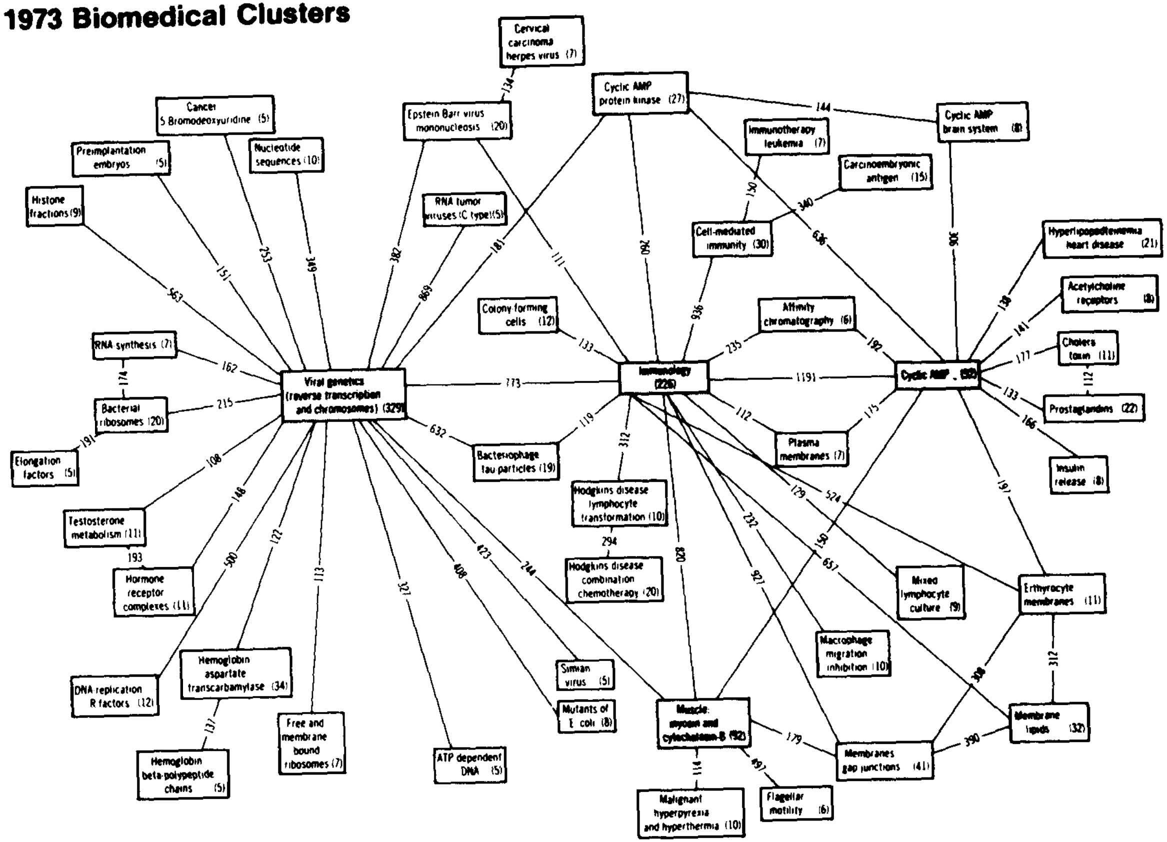

Figure 8.17 Map of 1973 biomedical clusters. Boxes represent clusters. Parenthetical numbers in boxes indicate the number of cluster documents. Co-citation links between clusters are shown by connecting lines; strength of links is shown by numbers in lines.

and *chromosomes* (*5* and *34*), *immunology* (*27*), *biological membranes* (*39, 40, 41*), and *cyclic AMP: protein kinase* and *microtubule protein* (*35*). By 1973 (Figure 8.17) *chromosomes; RNA viruses, reverse transcriptase;* and several smaller specialties concerned with the genetics of viruses have coalesced into *viral genetics,* which dominates the map. *Immunology* has grown and established a strong link with *viral genetics.* A direct, very strong link also has evolved between *immunology* and *cyclic AMP*; and *microtubule protein* has been transformed into *muscle, myosin and cytochalasin-B,* which is not only one of the four major specialties but which also is linked to each of the other three.

Small also tested the logic of the biomedical structure at the more abstract level of disciplines. Using the 1973 *SCI* data base of 2.7 million reference citations, he selected those that met a citation frequency threshold of 15, clustered them at a co-citation level of 7, and calculated the co-citation links between clusters. The result is the map of five natural science disciplines shown in Figure 8.18. It agrees reasonably well with the major features of the specialty map of the natural sciences produced from the *SCI* data on the first quarter of 1972 (Figure 8.10). *Biomedicine, chemistry,* and *physics,* which were the three major poles of the specialty map, account for three of the five disciplines identified. Also, as on the specialty map, *chemistry* is the only one of the three that is strongly linked to almost all areas, showing up again as the integration point for the natural sciences. Besides these similarities, other interesting features are that at the disciplinary level of abstraction, *physics* is still distinctly multipolar and *crystallography,* through strongly linked to *chemistry,* shows up as a separate discipline.

The most intensive validation effort made to date has involved detailed studies of a number of specialties over a period of time. Small looked at the behavior of 30 dif-

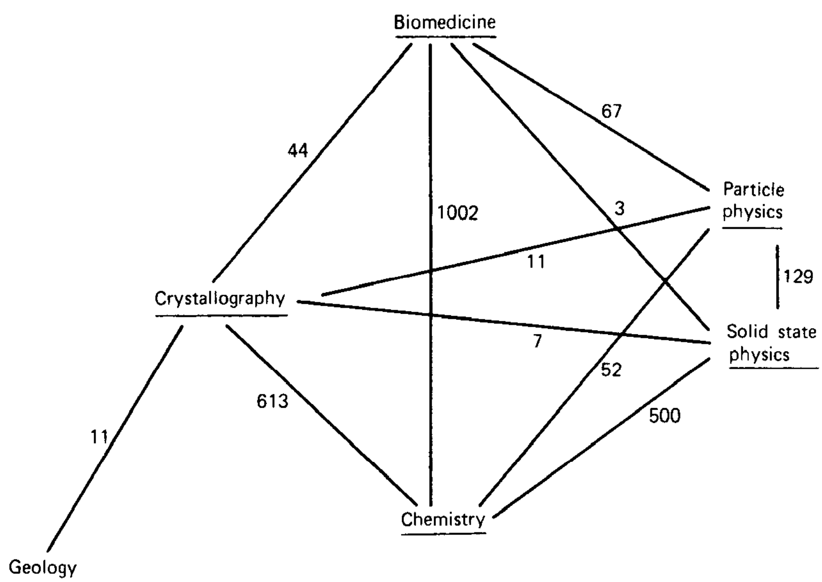

Figure 8.18 Map of major natural science disciplinary clusters in 1973. Numbers beside connecting lines indicate strength of co-citation links.

ferent specialty clusters during the four years of 1970–1973. Though his analysis is still incomplete, there are some interesting initial observations. One is cluster behavior that is consistent with the findings of a number of sociometric and historical studies (18, 19) that specialties can change or emerge very suddenly—in biomedicine, within six months after the publication of a keystone discovery paper (20). Also interesting is the finding that the average continuity rate of documents over the four-year period was 40%, which suggests that the intellectual fermentation of science proceeds at a slow boil. That figure, however, is only the average for the 30 specialties; the range of continuity rates is quite broad. In one-third of the specialties, all but one or two of the documents in the first-year cluster were gone by the fourth year. If high turnover rate (of documents) can be correlated with the revolutionary shifts of research focus theorized by Kuhn (21), this finding suggests that specialties average one revolution every 12 years. Another interesting facet of the change process is that documents tended to move in and out of the clusters in multiples, rather than one at a time. Change on this scale might be a sign that the leadership within the specialty has shifted from one group of researchers to another.

A much more complete picture of cluster behavior over time, and how that behavior matches the real-world events within the specialty, was produced by Small from a study of the literature on collagen research during the five-year period running from 1970 through 1974 (14).

Collagen is a large, triple-helical, protein molecule in the human body whose functional role is to form fibers that attach to tendons and joints and line arterial walls. The primary constituent of connective tissue, it is the most abundant protein in the body. Despite its ubiquity, collagen had the general reputation, for many years, of being a rather unexciting research subject. Until the 1960s, the consensus was that collagen was an inert material, and all the research was concerned with defining its structural characteristics. Initially, beginning in the 1920s, electron diffraction was the primary methodology used in the structural studies. In the 1950s, X-ray diffraction considerably increased the power of structural analysis and produced a molecular picture of collagen as being a triple-stranded, helical coil. Two of the strands, named alpha-1, were biochemically identical. The third was different, so was named alpha-2.

Most of the work in the 1960s was devoted to defining the structures of the strands. During this period, there was a switch to biochemical methodologies to work out amino acid sequences and to define the nature and composition of the linkages connecting the strands. The carboxymethyl cellulose chromatography column (CMC column) was used to separate the molecular chains, and cyanogen bromide (CNBr) was used to divide the chain into appropriate subunits.

A discovery, in 1969, of a genetically distinctive type of collagen that consisted of only the two alpha-1 strands set off a search for other types of collagen. Work during 1970 and 1971 confirmed a theory that predicted the existence of a soluble, precursor type of collagen, which eventually was named "procollagen." This discovery produced a radical shift in the research orientation of the specialty. Though the structural studies continued, the main focus of collagen research shifted to understanding the stability and biosynthesis of the collagen molecule.

The intellectual shift was accompanied by a sociological one. What had been considered a dull specialty suddenly became an exciting one. What had been a small, very cohesive area of biochemical research began growing in size and scope, generating subspecialities and attracting MD's working on the clinical medicine problems of aging, heart and lung disease, cancer, and arthritis. The cooperation that is often so typical of small specialties was replaced by the competition that is often typical of the larger ones.

That is the history of collagen as gleaned from its literature and conversations with its researchers. The version of the most recent part of that history that was derived from citation analysis is shown in Figures 8.19 through 8.23. The clusters, which were generated from the *SCI* data base at a citation-frequency threshold of 15, and a co-citation threshold of 11, are shown as contour maps. The documents of a cluster are represented by dots, whose relative locations have been plotted, by a metric-scaling algorithm, according to co-citation frequencies. In general, the proximity of each dot to all the rest is inversely proportional to their co-citation strength (the stronger the co-citation link, the closer the proximity). The contour lines are drawn to provide a height scale that is directly proportional to the citation frequency of each document. A bibliography of all the papers shown in the series of cluster maps appears in Figure 8.24.

The 1970 cluster (Figure 8.19) reflects a research orientation that is exclusively structural. Piez 1963, the most highly cited document, defines the structure of the

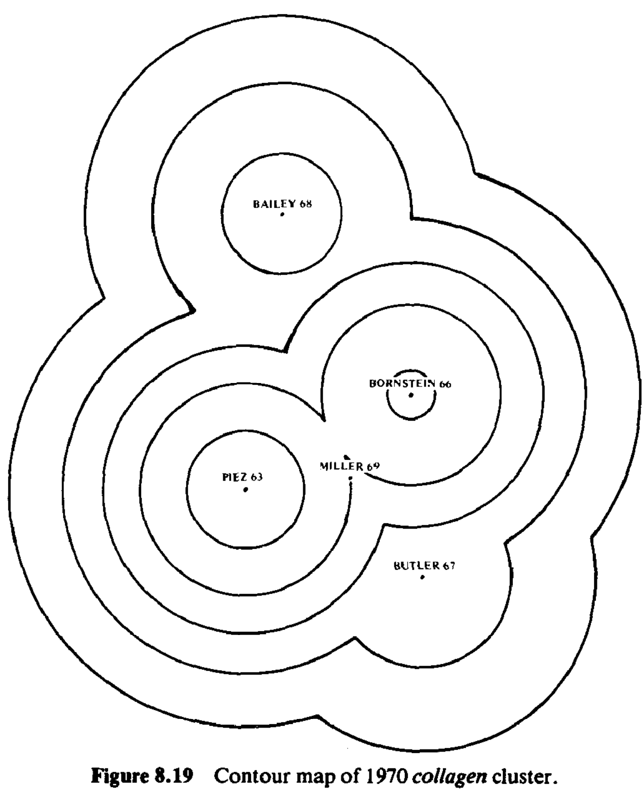

Figure 8.19 Contour map of 1970 *collagen* cluster.

alpha-1 chain. Bornstein 1966, the second most highly cited document, describes the nature of the links between chains. Butler 1967 and Miller 1969, whose proximity to each other reflects frequent co-citation, both discussed the same general aspects of the alpha-1 structure—the primary difference between them being that one used bone collagen for his studies whereas the other used skin collagen. Bailey 1968 also reports on a structural study, one that produced a catalog of the amino acids on the collagen cross-links. This document is set apart somewhat from the rest because the work it describes differs enough from the rest to make co-citation with the other documents relatively infrequent.

The 1971 cluster (Figure 8.20) is identical to that of 1970, except for the disappearance of Bailey 1968. The research continues to be exclusively structural.

The 1972 cluster (Figure 8.21) shows a major change. Suddenly there are two distinctly separate areas of collagen research. One is the structural research from past years, represented in the right-hand region by the documents from the 1971 cluster, plus one new entry: Rauterberg 1971, which describes work similar to that done by Butler 1967 and Miller 1969, except that the collagen analyzed is from a different source. The other area of research, represented by a completely new set of papers in the left-hand region, is concerned with the biosynthesis of collagen. This area is dominated by Layman 1971 and Bellamy 1971. Layman 1971 was the first to report the discovery of the new type of collagen that was later to be named procollagen. Bellamy 1971 was the one who named it, and the first to characterize it as the precursor to collagen. Müller 1971 was the first to determine whether the synthesis of collagen from procollagen took place inside or outside the cell. Dehm 1971

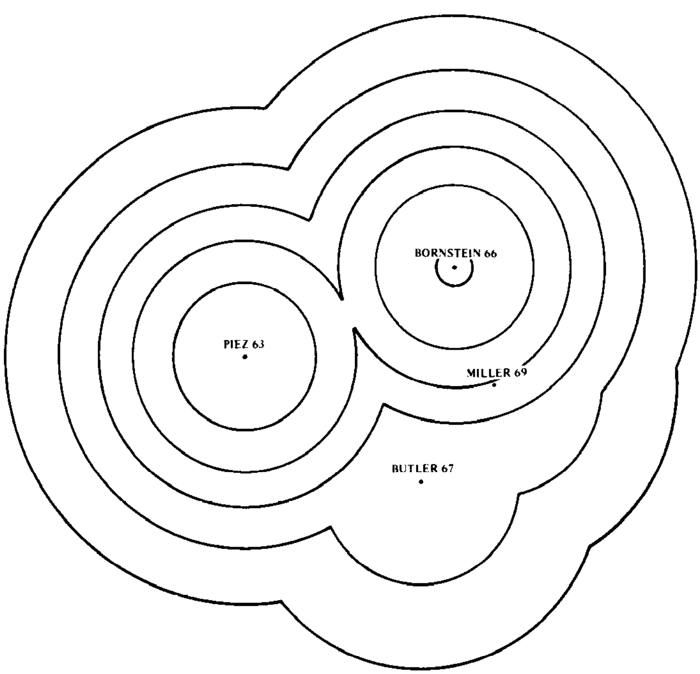

Figure 8.20 Contour map of 1971 *collagen* cluster.

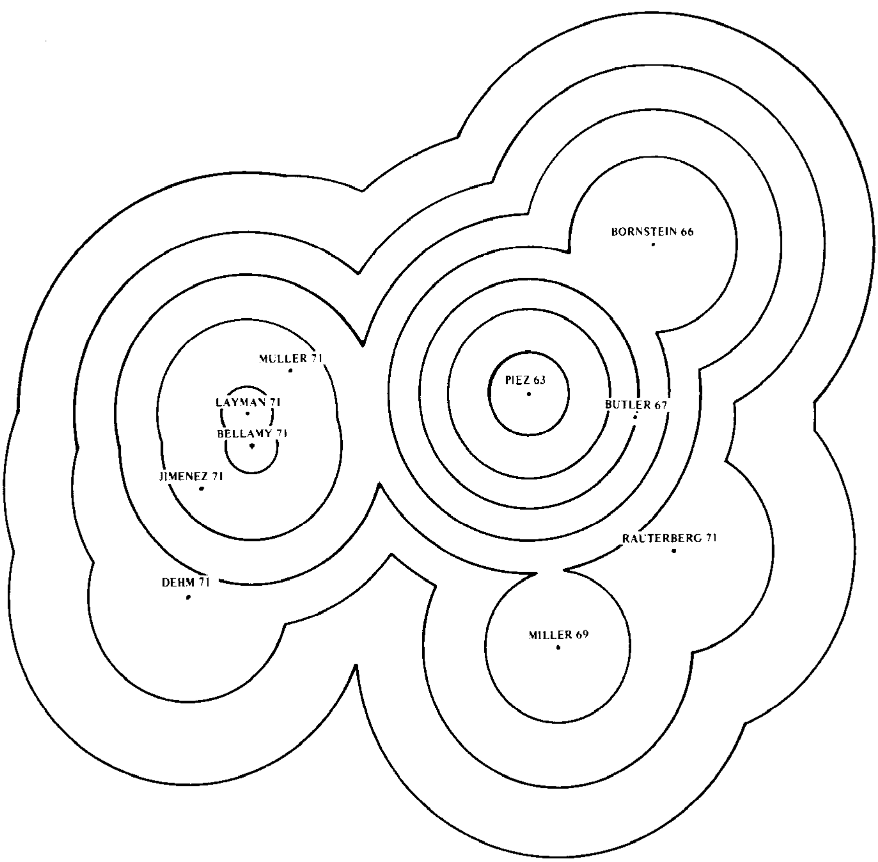

Figure 8.21 Contour map of 1972 *collagen* cluster.

described the preparation of a matrix of free cells that plays an important methodological role in biosynthesis studies. Jimenez 1971 used this methodology to determine the molecular size of procollagen.

An interesting characteristic of the series of new documents on biosynthesis is that they are all only one year old. The achievement of a high citation rate within a year after publication may be a sign of a discovery paper that has enough intellectual energy to change the state of a specialty.

The 1973 cluster (Figure 8.22) shows a return to a single research orientation, but that orientation now is biosynthesis. All the structural work has disappeared. The original work on procollagen has been retained, and new documents on the subject have appeared. Layman 1971 now clearly dominates the cluster, with Bellamy 1971 close by. Two new papers—Bornstein 1972 and Dehm 1972—have pushed Møller 1971 to the periphery, with studies on the cystine content of procollagen. A second group of new papers to the right of Layman, consisting of Stark 1971, Lenaers 1971, and Lapiere 1971, deal with the investigations of a Belgian laboratory into a cattle disease suspected of being caused by procollagen. Again, the currency of the cluster is still remarkable. None of the documents is more than two years old.

The 1974 cluster (Figure 8.23) shows a major increase in the size of the specialty.

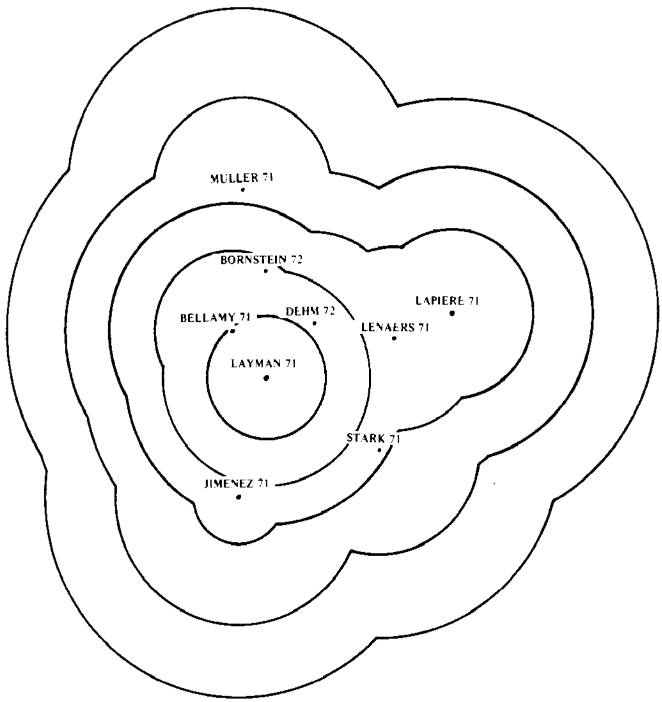

Figure 8.22 Contour map of 1973 *collagen* cluster.

The cluster is now almost three times its 1973 size, with the number of documents having increased from 9 to 24, and it identifies three relatively separate research fronts.

The group of five documents at the lower right are on the stabilization of collagen by hydroxyproline. Built on the existence and characterization of procollagen, this work represents the first major advance made possible by the original discovery of procollagen. The fact that it came only two years after the original research is a sign of rapid research growth and change. Typically, four of the five papers are only one year old, and the fifth is only two.

The central region of the cluster contains the original work on procollagen. Still dominated by Layman 1971, it has been expanded by three new documents.

The group of papers to the left of Layman 1971 are on genetically distinctive types of collagen—work that until this point had formed a cluster separate from the main line of collagen research. (This cluster was easily identifiable as a separate branch of work on collagen as early as 1972, but no one thought to look for more than one collagen cluster.) It is dominated by Miller, who discovered, in 1969, a form of collagen composed of only the two alpha-1 strands. This discovery set off the search for other types of collagen that culminated in the identification of procollagen. The appearance of this work in the main collagen cluster suggests that the relationship between research on genetic variations of collagen and the characterization of procollagen became considerably stronger in 1974.

Two papers in the 1974 cluster are somewhat removed from all three research

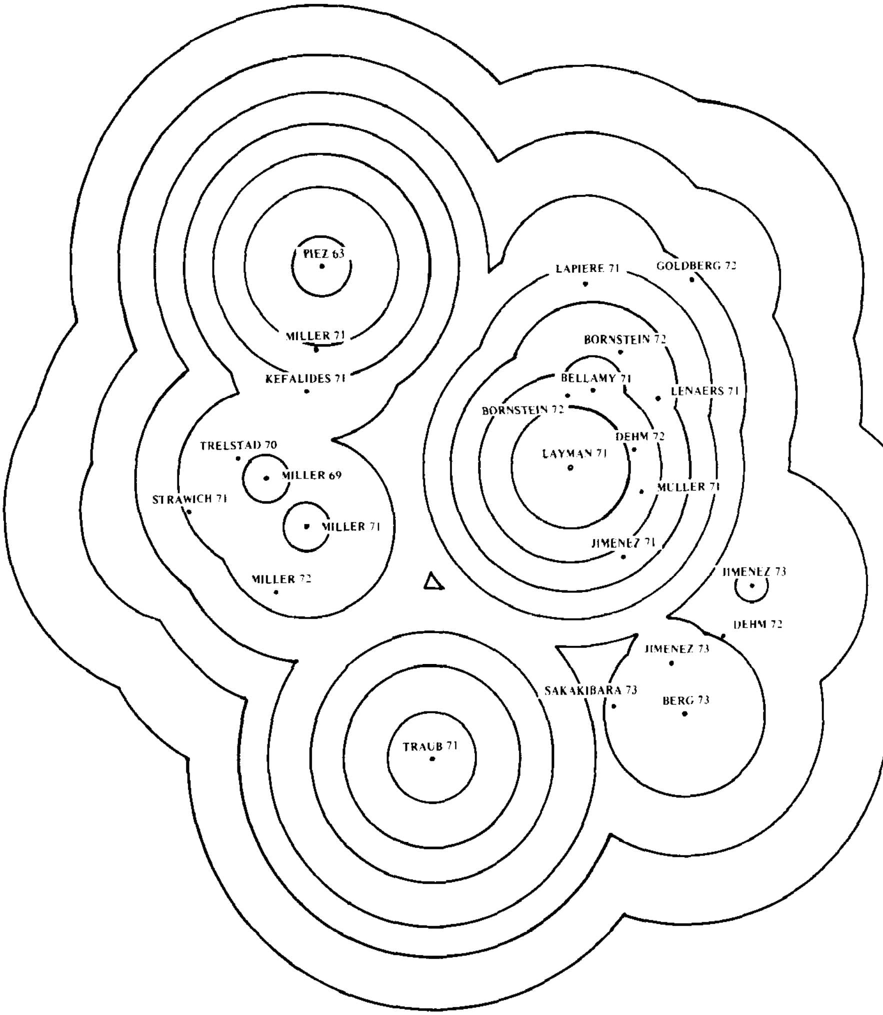

Figure 8.23 Contour map of 1974 *collagen* cluster.

fronts. The reason, it turns out, is not that they are unrelated to any of the fronts, but that they are related to all of them, rather than to any single one. Piez 1963 is the same document that was so prominent in the 1970–1971 clusters and then disappeared in 1973, when the research emphasis shifted from structural studies to biosynthesis. A landmark paper on structure, it has reappeared as a highly cited paper in 1974 because it is methodologically important to all the research fronts. Traub 1971 is a new paper that borders on all three of the fronts because it is a comprehensive review of the state of the art of the entire specialty.

Figure 8.24 Bibliography of documents in *collagen* cluster from 1970–1974.

Bailey, A. J. and **Peach, C. M.** "Isolation and Structural Identification of a Labile Intermolecular Crosslink in Collagen." *Biochemical and Biophysical Research Communications,* 33(5):812-819, 1968.

Bellamy, G. and **Bornstein, P.** "Evidence for Procollagen, a Biosynthetic Precursor of Collagen." *Proc. Nat. Acad. Sci. USA,* **68**(6):1138-1142, 1971.

Berg, R. A. and **Prockop, D. J.** "The Thermal Transition of a Non-hydroxylated Form of Collagen—Evidence for a Role for Hydroxyproline in Stabilizing the Triple-helix of Collagen." *Biochemical and Biophysical Research Communications,* **52**(1):115-120, 1973.

Bornstein, P. and **Piez, K. A.** "The Nature of the Intramolecular Cross-links in Collagen—the Separation and Characterization of Peptides from the Cross-link Region of Rat Skin Collagen." *Biochemistry,* **5**(11):3460-3473, 1966.

Bornstein, P., Ehrlich, H. P. and **Wyke, A. W.** "Procollagen: Conversion of the Precursor to Collagen by a Neutral Protease." *Science,* **175**:544-546, 1972.

Bornstein, P., Von Der Mark, K., Wyke, A. W., Ehrlich, H. P., and **Monson, J. M.** "Characterization of the Pro-q1 Chain of Procollagen." *Journal of Biological Chemistry,* **247** (9):2808-2813, 1972.

Butler, W. T., Piez, K. A. and **Bornstein, P.** "Isolation and Characterization of the Cyanogen Bromide Peptides from the q1 chain of Rat Skin Collagen." *Biochemistry,* **6**(12):3771-3780, 1967.

Dehm, P. and **Prockop, D. J.** "Synthesis and Extrusion of Collagen by Freshly Isolated Cells from Chick Embryo Tendon." *Biochimica et Biophysica Acta,* **240**:358-369, 1971.

Dehm, P., Jimenez, S.A., Olsen, B.R., and **Prockop, D.J.** "A Transport Form of Collagen from Embryonic Tendon: Electron Microscopic Demonstration of an NH_2-terminal Extension and Evidence Suggesting the Presence of Cystine in the Molecule." *Proc. Nat. Acad. Sci. USA,* **69**(1):60-64, 1972.

Dehm, P. and **Prockop, D.J.** "Time Lag in the Secretion of Collagen by Matrix-Free Tendon Cells and Inhibition of the Secretory Process by Colchicine and Vinblastine." *Biochimica et Biophysica Acta,* **264**:375-382, 1972.

Goldberg, B., Epstein, E.H., Jr., and **Sherr, C.J.** "Precursors of Collagen Secreted by Cultured Human Fibroblasts." *Proc. Nat. Acad. Sci. USA,* **69**(12):3655-3659, 1972.

Jimenez, S.A., Dehm, P., and **Prockop, D.J.** "Further Evidence for a Transport Form of Collagen—Its Extrusion and Extracellular Conversion to Tropocollagen in Embryonic Tendon." *FEBS Letters,* **17**(2):245-248, 1971.

Jimenez, S.A., Dehm, P., Olsen, B.R., and **Prockop, D.J.** "Intracellular Collagen and Protocollagen From Embryonic Tendon Cells." *Journal of Biological Chemistry,* **248**(2):720-729, 1973.

Jimenez, S.A., Harsch, M., and **Rosenbloom, J.** "Hydroxyproline Stabilizes the Triple Helix of Chick Tendon." *Biochemical and Biophysical Research Communications,* **52**(1):106-114, 1973.

Kefalides, N. A. "Isolation of a Collagen from Basement Membranes Containing

Figure 8.24 (continued)

Three Identical q-chains." *Biochemical and Biophysical Research Communications,* **45**(1): 226-234, 1971.

Lapiere, C.M., Lenaers, A. and **Kohn, L.D.** "Procollagen Peptidase: an Enzyme Excising the Coordination Peptides of Procollagen." *Proc. Nat. Acad. Sci. USA,* **68**(12): 3054-3058, 1971.

Layman, D.L., McGoodwin, E.B. and **Martin, G.R.** "The Nature of the Collagen Synthesized by Cultured Human Fibroblasts." *Proc. Nat. Acad. Sci. USA,* **68**(2): 454-458, 1971.

Lenaers, A., Ansay, M., Nusgens, B.V., and **Lapiere, C.M.** "Collagen Made of Extended q-chains, Procollagen, in Genetically Defective Dermatosparaxic Calves." *Eur. J. Biochem.,* **23**: 533-543, 1971.

Miller, E.J., Lane, J.M., and **Piez, K.A.** "Isolation and Characterization of the Peptides Derived from the q1 Chain of Chick Bone Collagen after Cyanogen Bromide Cleavage." *Biochemistry,* **8**(1): 30-39, 1969.

Miller, E.J. and **Matukas, V.J.** "Chick Cartilage Collagen: a New Type of q1 Chain Not Present in Bone or Skin of the Species." *Proc. Nat. Acad. Sci. USA,* **64**: 1264-1268, 1969.

Miller, J. "Isolation and Characterization of a Collagen from Chick Cartilage Containing Three Identical q Chains." *Biochemistry,* **10**(9): 1652-1659, 1971.

Miller, E.J., Epstein, E.H., Jr., and **Piez, K.A.** "Identification of Three Genetically Distinct Collagens by Cyanogen Bromide Cleavage of Insoluble Human Skin and Cartilage Collagen." *Biochemical and Biophysical Research Communications,* **42**(6): 1024-1029, 1971.

Miller, E.J. "Structural Studies on Cartilage Collagen Employing Limited Cleavage and Solubilization with Pepsin." *Biochemistry,* **11**(26): 4903-4909, 1972.

Müller, P.K., McGoodwin, E., and **Martin, G.R.** "Studies on protocollagen: Identification of a Precursor of Proto q1." *Biochemical and Biophysical Research Communications,* **44**(1): 110-117, 1971.

Piez, K.A., Eigner, E.A., and **Lewis, M.S.** "The Chromatographic Separation and Amino Acid Composition of the Subunits of Several Collagens." *Biochemistry,* **2**: 58-66, 1963.

Rauterberg, R. and **Kühn, K.** "Acid Soluble Calf Skin Collagen—Characterization of the Peptides Obtained by Cyanogen Bromide Cleavage of its q1-chain." *Eur. J. Biochem.,* **19**: 398-407, 1971.

Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and **Prockop, D.J.** "Synthesis of (Pro-Hyp-Gly)$_n$ of Defined Molecular Weights—Evidence for the Stabilization of Collagen Triple Helix by Hydroxypyroline." *Biochimica et Biophysica Acta,* **303**: 198-202, 1973.

Stark, M., Lenaers, A., Lapiere, C., and **Kühn, K.** "Electronoptical Studies of Procollagen from the Skin of Dermatosparaxic Calves." *FEBS Letters,* **18**(2): 225-227, 1971.

Strawich, E. and **Nimni, M.E.** "Properties of a Collagen Molecule Containing Three Identical Components Extracted from Bovine Articular Cartilage." *Biochemistry,* **10**(21): 3905-3911, 1971.

Traub, W. and **Piez, K.A.** "The Chemistry and Structure of Collagen." *Adv. Protein Chem.,* **25**: 243-352, 1971.

Trelstad, R.L., Kang, A.H., Igarashi, S., And Gross, J. "Isolation of Two Distinct Collagens from Chick Cartilage." *Biochemistry,* **9**(25): 4993-4998, 1970.

Figure 8.24 (continued)

One abstract, but revealing, quantitative measure of the five years of activity shown by the collagen clusters is a stability index, which is computed by dividing the number of documents that survive *n* years by the number of unique documents that appear in the cluster over the same period of time. The stability index for the collagen clusters over two- and three-year periods is shown in Figure 8.25.

	70-71	71-72	72-73	73-74
Stability Index	.80	.31	.32	.44

	70-71-72	71-72-73	72-73-74
Stability Index	.29	0	.20

Figure 8.25 Stability index for *collagen* cluster from 1970–1974.

Since the major conceptual shift in collagen research first showed up in the third year (1972) of the five years studied, the stability index is considerably lower for the successive three-year periods than it is for the two-year ones. In the 1971–1973 period, which brackets the emergence of the biosynthesis research front, the stability index is zero; not one of the papers in the 1971 cluster survived the two following years. In the 1972–74 period, the stability index rebounded to only .20, despite the fact that the biosynthesis front had emerged in 1972. In other words, even after the appearance of the discovery papers that provided the foundation for the new conceptual direction, the combination of turnover and net increase in core documents kept the stability index low, which suggests that the new research front remained very dynamic.

The same picture is shown, though not as dramatically, by the stability-index measures of successive two-year periods. It dropped from a high of .80 during 1970–1971, before the biosynthesis research was published, to a low of .31 during the 1971–1972 period, which included the last year in which all the work was structural and the first year in which biosynthesis emerged as an important research front. Continuity was barely any stronger during 1972–1973, when the structural work disappeared completely from the cluster, and rose to only .44, well below the presynthesis days, during the last two years.

PEER VALIDATION

Both a qualitative and quantitative analysis of the collagen cluster, then, characterized the five-year period as one in which the intellectual focus of the specialty changed completely, producing a major increase in size and development rate. This picture seemed to fit the general history of the specialty that had been put together from a review of the literature and conversations with several collagen researchers. But to test it more precisely, and in more detail, Small (14) surveyed 24 collagen researchers who had published papers in 1973 that cited at least one of the documents in the 1973 cluster.

The survey was conducted by mail. The researchers surveyed were not told anything at all about the citation analysis that had been performed, or about any of the data produced by the analysis. They were simply asked to respond to a few questions about the recent history of collagen research. The only impact that the citation analysis had upon the survey—and it was both subtle and indirect—was that it provided the underlying hypothesis that there was a specialty of collagen research. Eleven of the researchers surveyed responded completely, which confirmed that the specialty did exist and was properly named.

The survey, which was conducted at the end of 1974, asked four questions:

1. What were the most important scientific advances in collagen research during the past five years?
2. What papers were the first to describe the advances identified in question 1?
3. Whom do you regard as the leading investigator in collagen research?
4. Has collagen research undergone a major change or conceptual shift in the last five years? Please explain.

The responses to these questions are summarized in Figures 8.26 through 8.29.

In the first question, the respondents identified five advances that they considered important (Figure 8.26). All of them, except for the work on genetic defects, were identified by the citation analysis. The work on the sequencing of collagen chains was done prior to 1970, which put it outside the time frame of the study, but still was identified by papers in the 1970 cluster. The two advances that were most visible in the citation analysis—the discoveries of procollagen and genetically distinctive types of collagen—also turned out to be the only ones on which there was substantial agreement among the researchers.

Advance or Development	Number of Responses
1. Procollagen, a biosynthetic precursor molecule	11(100%)
2. Genetic types of collagen	10
3. Sequencing of collagen chains	5
4. Genetic defects of collagen metabolism in hereditary diseases	5
5. Stabilization of collagen by hydroxyproline	4

Figure 8.26 Responses to question about the most important advances in collagen research.

When it came to identifying the papers that first reported these advances, the respondents singled out 32 papers, 14 of which appeared in at least one cluster.

However, only nine of the 32 papers were mentioned by at least two respondents. Shown in Figure 8.27, along with the annual clusters in which they appeared, all but one of those nine papers were picked up by the citation analysis. The missing paper was Pinnell 1972, which described genetic defects of collagen metabolism—the one advance missed by the citation analysis. Again, as in the case of the advances, those papers about which there was the most agreement were all identified by the citation analysis.

Paper (First author, Year)	Mentions	Years in Cluster 70	71	72	73	74	Subject
1. Layman 1971	7			X	X	X	Procollagen
2. Bellamy 1971	7			X	X	X	Procollagen
3. Miller 1969	6			X	X	X	Genetic types
4. Jimenez 1973	3					X	Hydroxyproline
5. Berg 1973	3					X	Hydroxyproline
6. Jimenez 1971	2			X	X	X	Procollagen
7. Miller 1971	2			X		X	Genetic types
8. Kefalides 1971	2					X	Genetic types
9. Pinnell 1972	2						Hereditary disease

Figure 8. 27 Responses to question about the papers that reported the most important advances in collagen research. Twenty-three other papers received one mention. Six of these appeared in the cluster.

In checking back to determine why the Pinnell work was missed, Small found that the first year it could have been included (1973), it made the citation-frequency threshold of 15, but did not make the co-citation threshold of 11. The next year (1974), the Pinnell paper was cited fewer than 15 times, so it did not qualify for the data base subset from which document pairs were extracted for clustering. The threshold levels set obviously are very critical to how comprehensive a picture citation analysis can produce of the activity in a specialty.

When asked to identify the leading investigators in collagen research, the respondents identified 24 people, 15 of whom appeared in the clusters as authors or coauthors of at least one paper. Figure 8.28 shows the ones who were mentioned at least twice and the number of cluster papers they were associated with. Ten of the 13 were associated with at least one paper, and there is a general correlation between the number of times they were mentioned by their peers and the number of cluster

Investigator	Number of Mentions	Number of Authorships on Clustered Papers
1. Bornstein	11	5
2. Miller	10	5
3. Prockop	9	7
4. Piez	9	6
5. Martin	9	2
6. Kuhn	7	2
7. Gross	7	1
8. Viez	3	0
9. Bailey	2	1
10. Rosenbloom	2	1
11. Trelstad	2	1
12. Fessler	2	0
13. Ballop	2	0

Figure 8. 28 Responses to question about the leading investigators in collagen research. Eleven other investigators received one mention. Five of these appeared in the cluster.

papers that identified them. Interestingly enough, however, the correlation does not hold for those who were mentioned by their peers less than twice. Four investigators who were named by their peers less than twice (two, in fact, were not mentioned at all) were involved in writing at least three cluster papers. One of those who was not mentioned at all was an author of five highly cited papers. This finding suggests that many of the cluster authors are young, upcoming scientists whose reputation among their peers has not yet caught up with their work. In this sense, citation measurements of researcher prominence may be a leading indicator of peer judgment. More evidence of this is found in the fact that the cluster papers identified roughly twice as many researchers as the respondents, and many of them were young, relatively junior (in terms of experience) coauthors.

The response to the last question of whether there had been a major conceptual shift in the specialty and, if so, what precipitated it, is summarized in Figure 8.29. Ten of the 11 respondents said that a major change had taken place. As for the reason for the change, the discovery of procollagen was credited nine times; the discovery of genetically different collagen, five times; and the work on genetic defects, twice. Again, the experts working in the field agreed with the citation analysis. They confirmed the shift shown by the clusters, and they attributed the shift to the two developments that changed the size and configuration of the clusters.

Yes: 10
No: 1

Reason for shift	Number of Responses
1. Concern with collagen biosynthesis via procollagen	9
2. Concern with genetically different collagens	5
3. Concern with genetic defects in collagen metabolism	2

Figure 8. 29 Responses to question about major conceptual shifts in collagen research.

Overall, the survey of collagen researchers produced a five-year history of the specialty that matched the one derived from citation data. This result strengthens the hypothesis that clusters formed by this particular type of citation analysis correspond to scientific specialties and that the cluster characteristics provide a usefully accurate picture of the intellectual nature of the specialty, the rate and direction of its evolution, and the number and identity of its key people.

EXAMINING THE SOCIAL SCIENCES

While the major concern of these initial studies has been the development and validation of the methodology, Small also has demonstrated the universality of the methodology by using it to produce a coherent, though still unproved, picture of the structure underlying the social sciences.

Citation analysis, of one type or another, is not new in studies of the social sciences. Jaspars and Ackermans have used the citation patterns between journals to determine how well social psychology performs as a link between psychology and sociology (22). Lin has used the same patterns to study the relationship between sociology journals and institutions (23). Clark (24) and Myers (25) have both used citation rates to measure the eminence of psychologists. Citation data also has played a role in studies of the dynamics of two specialties in sociology. S. Cole used it to define the changes in deviance (26), while J. Cole and H. Zuckerman examined the sociology of science (27). In addition, the INFROSS project at Bath University has used citation relationships between journals and age distributions of references in extensive studies of the information-transfer patterns and requirements of the social sciences (28).

Since all these studies have established a correlation between citation patterns and the reality of the social sciences, it was reasonable to assume that the co-citation clustering technique might very well produce a coherent picture of the structure of the social sciences. Nevertheless, the outcome of the Small study (13) was far from certain. What Small was attempting was quite different, in both kind and degree, from all the previous studies. His was the first to use citation data to generate a large-scale map of all the social science specialties, and the size of the sample he used was several orders of magnitude bigger than any previous one.

The sample used was the *SSCI* data base for the three years from 1972–1974—a total of approximately 1.2 million unique, cited documents. A citation-frequency threshold of 10 produced a set of 14,110 documents, from which 830,042 co-cited pairs were extracted for clustering. The primary clustering was done at four co-citation threshold levels: 11, 13, 16, and 20.

The methodology of this study differed from the preceding ones in a single, but significant respect: the measure of co-citation strength. Rather than using the simple measure of the number of times a pair of documents was co-cited, the measure was normalized to take into account the total citation rate of the pair. A co-citation strength of 11, then, indicates that the co-citation rate of a pair is 11% of its total citation rate. This standard was applied also to the co-citation threshold levels set for qualifying pairs of documents for clustering: the clustering levels of 11, 13, 16, and 20 that were used in the study represent a ratio of co-citations to citations, rather than a simple co-citation count. In other words, to qualify for clustering at level 11, a pair of documents had to have a co-citation-to-citation ratio of at least 11%.

This change was adopted for two reasons. First of all, computing co-citation strength as a function of the citation rate makes the co-citation threshold less restrictive, which permits fields of relatively low activity to become visible. Second, it represses the methodological superstructures that tend to aggregate groups of individual specialties into macroclusters. The reason for this effect is that the highly co-cited methodology papers are generally co-cited with a large number of partner documents, which means that the co-citation strength achieved with any one document is a relatively small part of its total citation rate. The result of these two effects is a greater number of clusters and a more realistic distribution of cluster sizes.

Using this more refined methodology, the study explored the social sciences at three different structural levels: that of individual specialties, groups of closely related specialties, and a multidisciplinary aggregate of all specialties. In addition, the basic structural characteristics of the social sciences were compared with those of the natural sciences.

For the comparative analysis, the three-year sample of 14,110 cited social science documents that qualified for pairing by meeting the citation-frequency threshold of 10 was compared to a one-year sample (1973 *SCI*) of 15,973 cited natural science documents that qualified for pairing by meeting a citation-frequency threshold of 15. The difference in the citation-frequency thresholds used was intended to keep the data bases from which pairs were extracted as close as possible in size.

The major difference seen between the two structures was that the social sciences were much more tightly integrated. This was shown by several measures, the major one being the ratio of actual to potential pairs. By this measure, the degree of integration for the social sciences was 83% versus 56% for the natural sciences.

Other measures of cohesiveness showed the same sort of difference. The mean number of co-citation linkages per document was 117.8 for the social sciences versus 89.7 for the natural sciences. The nodal number of linkages per document was 57 for the social sciences versus 35 for the natural sciences. The same pattern held when the co-citation links between clusters formed at level 16 were worked out: 3.4% of the potential linkages between social-science clusters at that level were realized versus 2% for the natural sciences, and the mean number of linkages per social science cluster was 41.2 versus 32 for the natural science clusters.

Aside from this evidence of a greater degree of interaction between social science specialties, the only other obvious sign of a significant difference between the social and natural sciences was in the age of the cluster documents. The mean date for the natural science documents was 1969.6 versus 1966.5 for the social sciences. This finding is consistent with the often-heard opinion that the social sciences change more slowly than the natural sciences and that the concepts underlying its research are older.

In all other ways, the structures of the two parts of science appear to be much the same. The most significant similarity is that the specialty area of research seems to be the basic structural unit in both.

One view of the structure of the social sciences at a specialty level of detail is shown in Figure 8.30. This map was produced by clustering the data base of document pairs at a co-citation threshold level of 16%, and then reclustering the clusters (rather than the documents in the clusters) that had been cited at least 100 times at co-citation threshold levels of 20% and 10%.

In a sense, then, the map is a composite. The large network of specialties connected by solid lines was produced when the level 16 clusters were reclustered at level 20. The four ovals around the perimeter are separate networks, formed at level 20, that are shown in macrocluster form to fit them onto the map. They connected to the large network at the positions shown by the dotted lines only when the reclustering level was dropped from 20 to 10.

Another bit of perspective that is useful in interpreting the map is the fact that the level 16 clustering of co-cited document pairs produced 143 clusters that had been cited at least 100 times. When these specialties were reclustered at level 20, 47 of them formed the solid line network, which consists of psychology (mostly social, but with some experimental) and sociology. The only other networks of significant size that were formed were the ones shown in ovals as macroclusters: a 20 cluster network on *memory and learning; a three cluster network on multidimensional scaling;* a five cluster network on *psychiatry*; and a three cluster network on *counseling,* which is really a secondary psychiatry network. Not shown is a network of three clusters on law.

On the one hand, the fact that all the networks link together at a reclustering level of 10 is a demonstration of the relatively high degree of integration in the social sciences. The level 10 network accounted for 70% of the 143 clusters originally created.

On the other hand, considering the high degree of integration, the fact that the *memory and learning* and *psychiatry* networks did not link to the main one until the co-citation threshold was dropped to 10 suggests that there are major conceptual and methodological differences between the specialties in these areas of research and the ones in social psychology and sociology. In retrospect, that does not seem very surprising in the case of *psychiatry.* However, it is surprising in the case of *memory and learning,* which is something one would expect to be in the mainstream of psychology.

The map also suggests that many of the social science specialties are highly problem oriented and employ a broad mix of scientific skills. For example, when the documents in the *organizational structure* cluster were examined, they were found to be a mix of material on psychology, sociology, and management science. Consistent with this characteristic, there are no specialties that can clearly be defined as social psychology. The type of work normally described by these names is scattered throughout the dominant psychology-sociology network.

What the map does not show, and its absence is somewhat conspicuous, is work in the areas of economics and political science. This gap was caused by the decision to recluster only those specialties whose documents had been cited at least 100 times. Characterized by low publication rates and low citation levels, economics and political science failed to qualify.

To provide a more comprehensive view of the social sciences structure, it was mapped a second time with the same technique used to map the natural sciences. Simply clustering the data base of co-cited pairs at a co-citation frequency threshold of 11% produced a map that was more inclusive and that struck a different balance between aggregation and differentiation. The result was a consolidation of most of the psychology clusters that dominated Figure 8.30 and the identification of specialties in economics, political science, psychoanalysis, and sociology that were missing from that map.

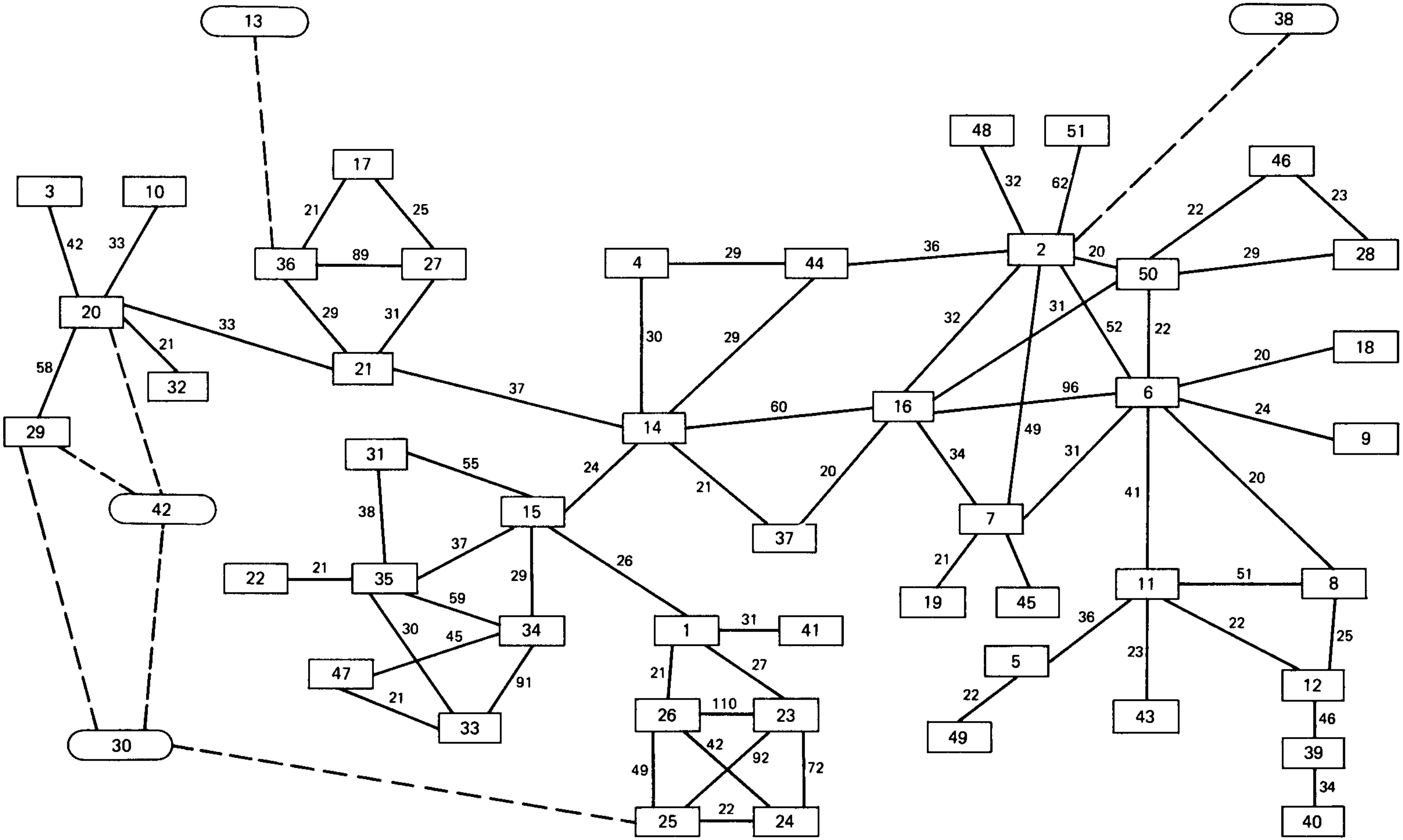

Figure 8.30. Map of major social sciences clusters from 1972–1974 data base.

KEY

1. Achievement motivation
2. Anxiety reduction
3. Attitude and behavior
4. Attitude change
5. Avoidance learning
6. Behavior modification
7. Behavior therapy
8. Behavioral contrast
9. Childhood psychosis
10. Cognitive balance
11. Concurrent reinforcement
12. Contrast in conditioning
13. Counseling
14. Equity theory
15. Expectancy theory predictions
16. Helping behavior
17. Human territoriality
18. Hyperactive children
19. Hypothalamic feeding
20. Impression formation
21. Interpersonal attraction
22. Leadership style
23. Locus of control
24. Locus of control, activism
25. Locus of control, alienation
26. Locus of Control, internal-external
27. Man-made environments
28. Mass-media violence
29. Measurement of human judgment
30. Memory and learning
31. Motivation and job satisfaction
32. Multiple-cue learning
33. Organizational decision making
34. Organizational structure
35. Organizational theory, management
36. Personal space
37. Prisoners dilemma
38. Psychiatry
39. Reinforcement in conditioning
40. Reward magnitude
41. Risky shift
42. Scaling, multidimensional
43. Schedule-induced behaviors
44. Self perception
45. Sexual behavior
46. Social aggression
47. Social participation in organization
48. State-trait model, anxiety
49. Taste-aversion learning
50. Television imitation
51. Treatment of phobias

The 20 largest clusters from the second, more inclusive map are shown in Figure 8.31, where they have been plotted as a series of points on a two-dimensional scale. The distance between points is inversely proportional to the strength of the co-citation links that join them (the closer the points, the stronger the co-citation links between them). The dominant cluster is *psychology,* which occupies the central position on the map. A number of separate psychology specialties, mostly clinical in nature, appear in the upper-right quadrant. Specialties in economics, political science, management, and sociology appear in the lower-left quadrant.

The most striking feature of the configuration of this map is the clear distinction it makes between the behavior of individuals and the behavior of groups. The specialties concerned with individual behavior are located in the upper-right quadrant, while those concerned with group behavior are in the lower left. This results in a well defined polar axis, similar to the one formed in the natural sciences by medicine, biology, chemistry, and physics, in that order. There is evidence that if co-citation clustering were used to produce a combined map of the social and natural sciences, the linear structure of the two would match, with the clinical psychology research on individuals (upper-right quadrant) leading into medicine.

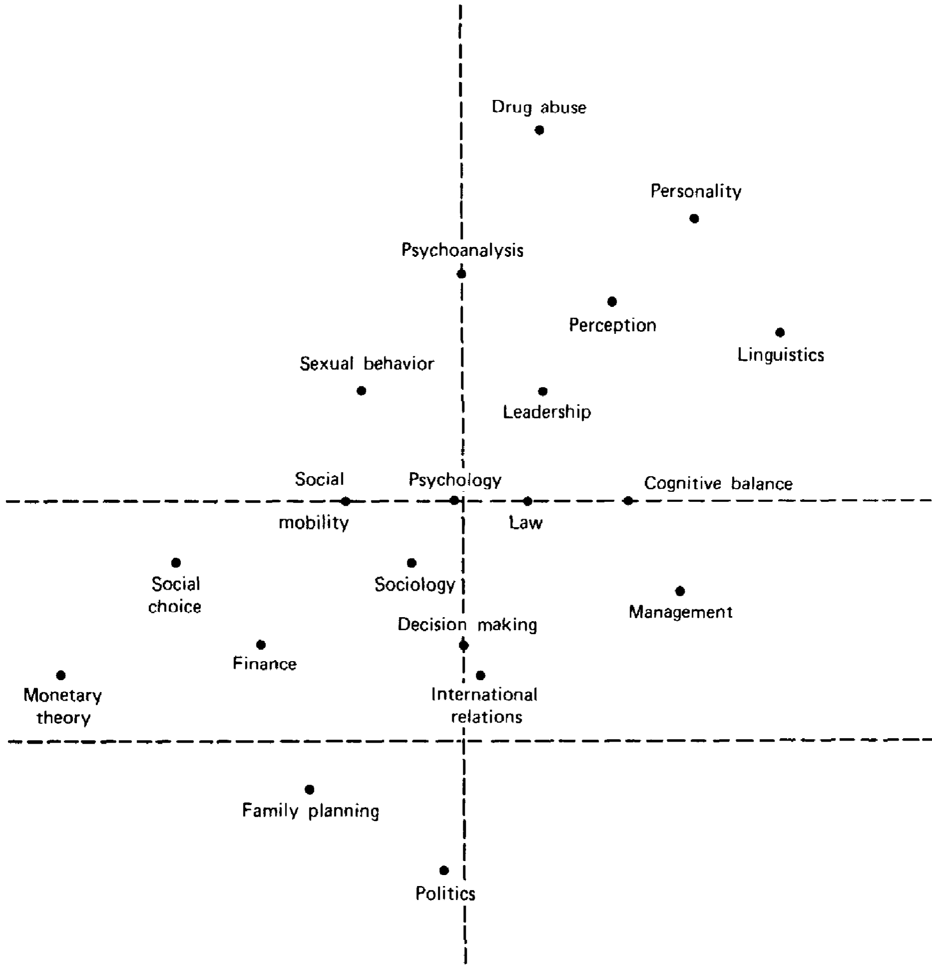

Figure 8.31 Two-dimensional plot of 20 largest social science clusters at level 11.

At the opposite end of the structural spectrum, the co-citation clustering methodology produced the same type of picture of individual social science specialties as it did of the ones in the natural sciences. Figure 8.32 shows an example: a specialty named *free recall,* which is one of the components of the *memory and learning* macrocluster on the social sciences map in Figure 8.30. Each box represents a document, which is described by the name of its author, the year and journal in which it was published, and the number of times it was cited. This shows the work on *free recall* as being dominated by five men: Craik, Glanzer, Waugh, Atkinson, and Rundus.

Figure 8.33 is an example of what was shown at an intermediate point on the structural scale. The subject is *memory and learning,* the macrocluster in which *free recall* is found. This view of the macrocluster was produced by identifying all the level 16 clusters that had at least 50 co-citation links to *free recall,* representing each one with a circle whose diameter is proportionately scaled to the number of times it was cited and then plotting them on a two-dimensional scale so that the overlap between circles is proportional to the strength of the co-citation links between clusters.

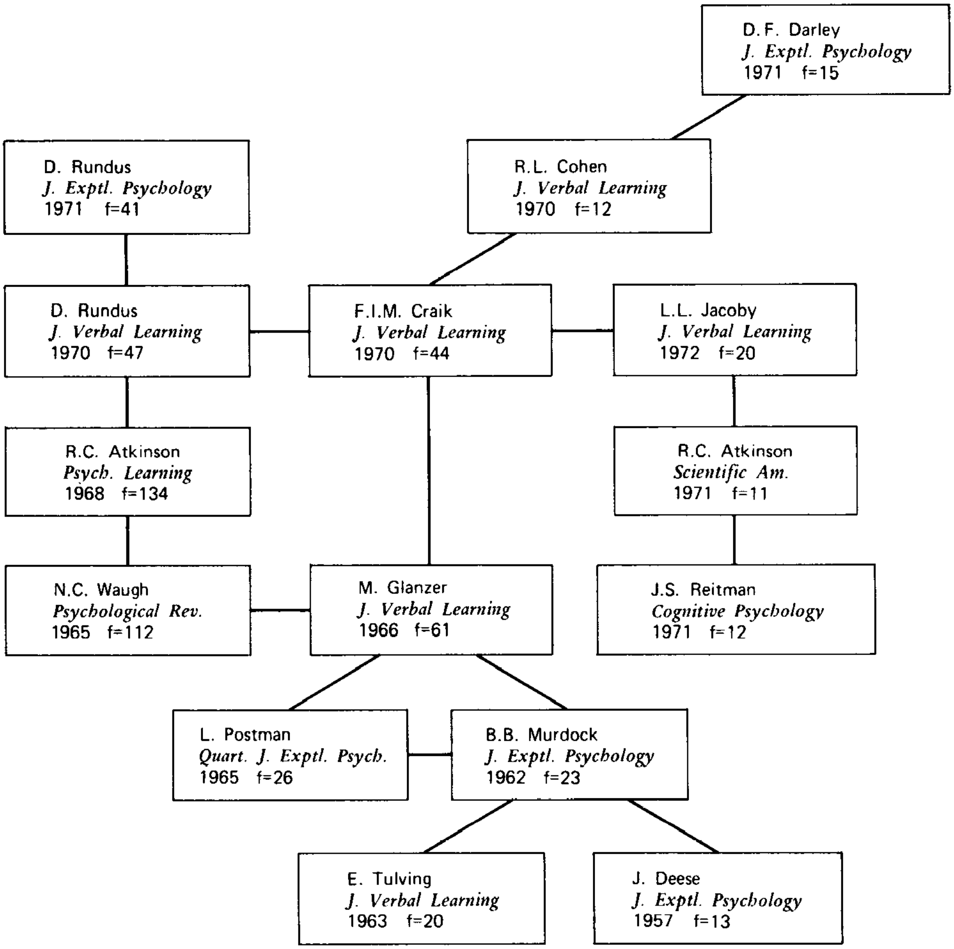

Figure 8.32 Block diagram of *free recall* cluster. Each block represents a document, which is described by the name of its author, the year and journal in which it was published, and the number of times it was cited from 1972–1974.

The result shows a very tightly structured set of 11 specialties in which *free recall* lies close to the center.

One of the interesting byproducts of this structural study of the social sciences is a list of the 26 documents that were cited most frequently during the 1972-1974 period (Figure 8.34). All of them were cited at least 200 times during the three-year period.

The most striking feature of this list is that all but three of the documents are books, which is consistent with the evidence and opinions that the ideas and concepts that support the research fronts in the social sciences are older than those found in the natural sciences. Another characteristic of the list that is consistent with the structural studies is the dominance of psychology. At least 15 of the 26 items on the list could be classified as belonging to that discipline. The overriding methodological importance of statistics in the social sciences is reflected by two texts on that subject being the two most frequently cited items on the list.

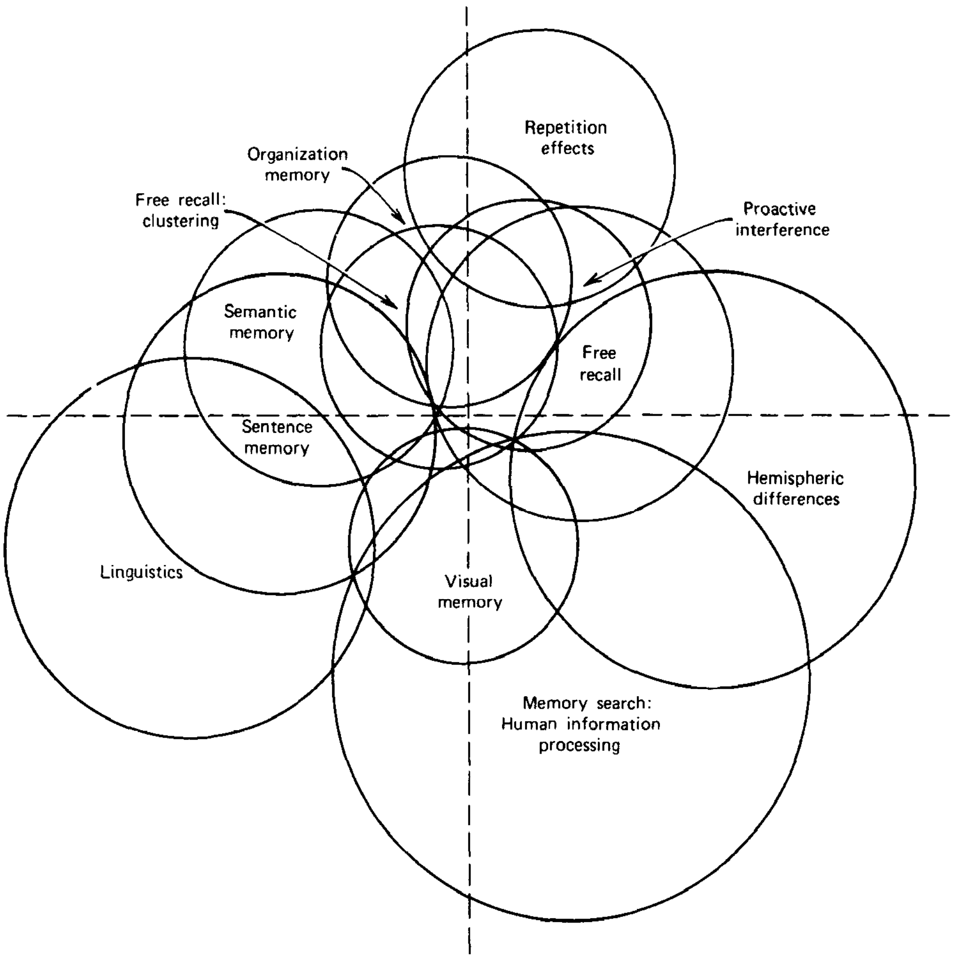

Figure 8.33 Spatial diagram of *memory and learning* macrocluster.

IMPLICATIONS

The ultimate implications of this series of studies are, of course, still uncertain, but they could be highly significant. Although Price has long hypothesized that science, despite its complexity, must be a two-dimensional spatial system (29), the co-citation maps of the natural and social sciences are the first confirmation of that hypothesis. Price, for one, thinks the confirmation has important theoretical and practical implications for scientific communication, particularly as it pertains to the number and subject scope of primary journals and to indexing systems (30). One very practical communications implication that ISI already is seriously working on is the development of an *Atlas of Science* that provides the same type of graphic, two-dimensional guide to the intellectual world of science as the traditional atlases do to the physical world (31).

Beyond that, the evidence produced by these initial studies suggests that the literature model of science may have led us to a methodology that will permit the

Author	Title	Year	Citation Frequency 1927–1974
Winer, B.J.	Statistical Principles in Experimental Design	62	1333
Siegel, S.	Nonparametric Statistics for the Behavioral Sciences	56	1020
Osgood, C.E.	Measurement of Meaning	57	480
Bandura, A.	Principles of Behavior Modification	69	422
Chomsky, N.	Aspects of the Theory Syntax	65	382
Rotter, J.B.	Psychological Monographs	66	381
Neisser, U.	Cognitive Psychology	67	334
Festinger, L.	Theory of Cognitive Dissonance	57	328
Heider, F.	Psychology of Interpersonal Relations	58	315
Merton, R.K.	Social Theory and Social Structure	57	313
Hollingshead, A.B.	Social Class and Mental Illness	58	303
Coleman, J.S.	Equality of Educational Opportunity	66	296
Hays, W.L.	Statistics for Psychologists	63	283
Rokeach, M.	Open and Closed Mind	60	277
Wolpe, J.	Psychotherapy by Reciprocal Inhibition	58	275
Toffler, A.	Future Stock	70	268
Adorno, T.W.	Authoritarian Personality	50	240
Jensen, A.R.	How Much Can We Boost IQ and Scholastic Achievement	69	226
Witkin, H.A.	Psychological Differentiation	62	219
Kuhn, T.S.	Structure of Scientific Revolutions	62	217
Thorndike, E.L.	Teachers' Word Book of Thirty Thousand Words	44	217
Parsons, T.	Social System	51	214
Skinner, B.F.	Beyond Freedom & Dignity	71	208
Miller, G.A.	The Magic number Seven: Plus or Minus Two	56	203
Katz, D.	Social Psychology of Organizations	66	202
Fenichel, O.	Psychoanalytic Theory of Neurosis	45	201

Figure 8. 34 List of social science documents cited at least 200 times from 1972 to 1974.

systematic investigation and objective definition, through structural analysis, of the status, dynamics, and underlying processes of science. But a lot of work is required to make that case conclusive.

There is no doubt that co-citation clustering has uncovered a structure in science. What is in doubt is the nature of that structure—whether it is the fundamental one of specialties that has been predicted, and is expected to produce informative insights. The studies conducted so far indicate that it is.

The collagen study proved that the cluster that appeared to represent the specialty of collagen research did, in fact, do so—at least to the extent that the changes that took place in its size, configuration, and intellectual content over time were accurate indicators of the state of the specialty as seen by the researchers who knew it best. But that study proves the validity of only one set of clusters. Similar success in a statistically significant number of specialty studies is needed to conclusively prove that all the clusters are representative of specialties.

Assuming that the clusters do represent specialties, the next step is to determine what types of information can be derived from them. The stability index, for example, must be validated as a measure of the rate of intellectual change in a specialty; and, if it is, it must also be calibrated so that we know what degree of change in the index is significant and the nature of the significance.

Furthermore, the emergence of a series of very recent papers describing a different research front must be validated as a signal of a major conceptual shift. And it must be determined whether there is a limit to the number of research fronts that can be contained in a single cluster before one of them is spun off to form a new independent specialty.

Cluster configurations must be studied to see if there is any relationship between the size and shape of clusters and the status of the specialty. It is possible that defining the status of science can be reduced to a pattern-recognition problem.

Another relationship that must be explored is the one between the intellectual structure of a specialty and its social structure. It is hard to imagine that the picture co-citation clusters produce of intellectual structures is completely divorced from social factors. But in what specific way are they related? Do the authors of the cluster documents represent the "invisible college" upon which the specialty is built? Does a change in these authors signal a change in the membership of that college? Is the size of the cluster related to whether communication within the specialty is informal and open or formal and guarded?

Once all these things are determined, cluster analysis can be employed, first, to test and further develop theories of specialty behavior and, then, to apply these theories to the job of characterizing the mosaic of science in ways that will be useful to the people who must make science policy decisions.

There is already much to test. At a general level, there are three hypothetical models of the dynamics of science that suggest particular patterns of specialty behavior. Kuhn thinks scientific change takes place in revolutionary spurts, interspersed with periods of stability (21). Popper sees it as proceeding at a continuous, rather than intermittent, revolutionary pace (32). And Toulmin sees the change as being continuous and evolutionary (33).

At the more detailed level, there are such things as Crane's observation that the growth in the size of a specialty seems to follow, rather than precede, innovation (4)—a pattern that certainly was seen in the collagen clusters, which did not expand dramatically in size until three years after the publication of the discovery papers that opened up a major, new research front. Another characteristic of specialty behavior that is important to study is the aging process. Is there a shift from basic to applied research as a specialty ages, or is there just a temporary increase in size and scope that is limited by a tendency for application work to achieve independent specialty status? And that question is but a part of a bigger, more important aspect of specialty behavior. Is there a universal life cycle for specialties—a cycle of birth, growth, and decay? If so, it would simplify the job of determining which of the three general models of science is most accurate.

In other words, there is still much validation, characterization, and application work to be done before anything definite can be said about the impact that co-citation clustering will have on the study and management of science. To facilitate the work, ISI is developing an on-line computer-based system that will permit researchers to work with the *SCI* and *SSCI* data bases in an interactive mode to produce and analyze diagrams of specialty clusters at all points on the structural scale. The storage-and-retrieval subsystem already is operational and has demonstrated a functional capability and efficiency that can considerably reduce the time and cost of performing citation analyses (34). The National Science Foundation is financing the development of the graphic sybsystem as part of their continuing interest in the potential of citation analysis as a science policy tool.

In his famous book, *The Structure of Scientific Revolutions* (21), Kuhn wrote,

"... 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 text-books 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."

Co-citation analysis promises to produce an index that does that and more.

REFERENCES

1. **Garfield, E., Sher, I. and Torpie, R.J.** *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 1964). 86 pp.
2. **Crawford, S.** "Informal Communication among Scientists in Sleep Research." *Journal of the American Society for Information Science,* **22**:301-310, 1971.
3. **Griffith, B.C. and Mullins, N.C.** "Coherent Social Groups in Scientific Change," *Science,* **177**:959-964, 1972.
4. **Crane, D.** *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago: University of Chicago Press, 1972). Pp. 2131
5. **Price, D.J.D.** Unpublished studies (New Haven: Yale University, 1964-1967).
6. **Price, D.J.D.** "Networks of Scientific Papers." *Science,* **149**:510-515, 1967.
7. **Narin, F., Carpenter, M. and Berlt, N.C.** "Inter-relationships of Scientific Journals." *Journal of the American Society for Information Science,* **23**:323-331, 1972; **Carpenter, M. and Narin, F.** "Clustering of Scientific Journals." (In press).

8. **Goffman, W.** "Mathematical Approach to the Spread of Scientific Ideas—The History of Mast Cell Research." *Nature,* **212**:449–452, 1966.
9. **Jahn, M.** *Changes with Growth of the Scientific Literature of Two Biomedical Specialties.* (Philadelphia: Drexel University, MS thesis, 1972). 64 pp.
10. **Small, H.** *Nuclear Physics in the Physical Review 1927-1940.* Unpublished report of a study at the American Institute of Physics, 1970.
11. **Small, H.G. and Griffith, B.C.** "The Structure of Scientific Literature, I: Identifying and Graphing Specialties." *Science Studies,* **4**:17–40, 1974.
12. **Griffith, B.C., Small, H.G., Stonehill, J.A., and Dey, S.** "The Structure of Scientific Literature, II: Toward a Macro- and Micro-structure for Science." *Science Studies,* **4**:339–365, 1974.
13. **Small, H.G.** *Mapping the Specialty Structure of the Social Sciences Using the Social Sciences Citation Index.* Unpublished report of a study for the National Science Foundation, 1975.
14. **Small, H.G.** "A Co-Citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research." *Social Studies of Science,* **7**:139–166, 1977.
15. **Kessler, M.M.** "Bibliographic Coupling Between Scientific Papers." *American Documentation,* **14**:10–25, 1963.
16. **Small, H.G.** "Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents." *Journal of the American Society for Information Science,* **24**:265–269, 1973.
17. **Marshakova, I.V.** "Bibliographic Coupling System Based on Cited References (*Science Citation Index*)." *Nauchno Tekhnicheskaya, Informatsiza Seriya,* **2**:3–8, 1973.
18. **Goffman, W.** "A Mathematical Method of Analyzing the Growth of a Scientific Discipline." *Journal of the Association for Computing Machinery,* **18**:173–185, 1971.
19. **Magyar, G.** "Bibliometric Analysis of a New Research Sub-field." *Journal of Documentation,* **30**:32–50, 1974.
20. **Griffith, B.C.** "On the Nature of Social Science and Its Literature." *CACUL Newsletter,* **4**:195–229, January 16, 1973.
21. **Kuhn, T.S.** *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962). Pp. 172
22. **Jaspars, J.M.F. and Ackermans, E.** "The Interdisciplinary Character of Social Psychology." *Sociologia Neerlandica,* **4**:62–79, 1966–1967.
23. **Lin, N.** "Stratification of the Formal Communication System in American Sociology." *American Sociologist,* **9**:199–206, 1974.
24. **Clark, K.E.** *America's Psychologists: A Survey of a Growing Profession* (Washington, D.C.: American Psychological Association, 1957). Pp. 247
25. **Myers, C.R.** "Journal Citations and Scientific Eminence in Contemporary Psychology." *American Psychologist,* **25**:1041–1048, 1970.
26. **Cole, S.** "The Growth of Scientific Knowledge: Theories of Deviance as a Case Study." In Coser, L.A. (ed.). *The Idea of Social Structure: Papers in Honor of Robert K. Merton* (New York: Harcourt, Brace, Jovanovich, 1975). Pp. 175–220.
27. **Cole, J.R. and Zuckerman, H.** "The Emergence of a Scientific Specialty: The Self-exemplifying Case of the Sociology of Science." In Coser, L.A. (ed.). *The Idea of Social Structure: Papers in Honor of Robert K. Merton* (New York: Harcourt, Brace, Jovanovich, 1975). Pp. 139–174.
28. Bath University Library. "The Use of Citation Linkages and Networks for information Retrieval in the Social Sciences." Working Paper #6, *Design of information Systems in the Social Sciences* (March 1973). See also Working Paper #11 on journal clustering.
29. **Price, D.J.D.** "Science: The Science of Scientists." *Medical Opinon and Review,* **1**:88–97, 1966.
30. **Price, D.J.D.** "Society's Needs in Scientific and Technical Information." *Ciencia da Informacao, Rio de Janeiro,* **3**:97–103, 1974.
31. **Garfield, E.** "ISI's Atlas of Science May Help Students in Choice of Career in Science." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977) 311–314.
32. **Popper K.** "Normal Science and Its Dangers." In Lakatos, I. and Musgrave, A. (eds.). *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press 1970). Pp. 51–58.

33. **Toulmin, S.** *Human Understanding: The Collective Use and Evolution of Concepts* (Princeton: Princeton University Press, 1972).

34. **Yermish, I.** "*A Citation-Based Interactive Associative Information Retrieval System.* (Philadelphia: University of Pennsylvania Ph.D dissertation, 1975). 278 pp., mimeogr.

To Contents Page

To Chapter 9