Author Cocitation: A Literature Measure of Intellectual Structure

Howard D. White and Belver C. Griffith
School of Library and Information Science, Drexel University, Philadelphia, PA 19104

It is shown that the mapping of a particular area of science, in this case information science, can be done using authors as units of analysis and the cocitations of pairs of authors as the variable that indicates their "distances" from each other. The analysis assumes that the more two authors are cited together, the closer the relationship between them. The raw data are cocitation counts drawn online from Social Scisearch (*Social Sciences Citation Index*) over the period 1972-1979. The resulting map shows (1) identifiable author groups (akin to "schools") of information science, (2) locations of these groups with respect to each other, (3) the degree of centrality and peripherality of authors within groups, (4) proximities of authors within group and across group boundaries ("border authors" who seem to connect various areas of research), and (5) positions of authors with respect to the map's axes, which were arbitrarily set spanning the most divergent groups in order to aid interpretation. Cocitation analysis of authors offers a new technique that might contribute to the understanding of intellectual structure in the sciences and possibly in other areas to the extent that those areas rely on serial publications. The technique establishes authors, as well as documents, as an effective unit in analyzing subject specialties.

Introduction

In the mapping of scientific literatures, Small [1], Small and Griffith [2], Griffith et al. [3], and recently many others have used documents as the unit of analysis and cocitations of pairs of documents as the variable that both indicates their "distance" from each other and permits their clustering into groups. Broadly, the more two documents are cited together, the closer the relationship between them, as perceived by the citing authors, and the closer they would appear in the graphic rendering of groups of documents. Treating documents as points and their cocitation levels as the inverse of distance creates maps. Such maps show the relationships between all pairs of documents at any chosen level: that of science as a whole, or of particular disciplines, specialties, or subspecialties.

The present article uses not individual documents as the units to be mapped, but *sets* of documents associated with the names of 39 information researchers, as first or single authors. It is, therefore, a cocited author analysis, mapping a single field and using as raw data the cocitation counts from Social Scisearch (*Social Sciences Citation Index* online) over the period 1972-1979. We are identifying and displaying a current picture of schools of research in information science, as represented by the authors that constitute them.

"Author" in this context means something like what the French call an *oeuvre*—a body of writings by a person—and not the person himself.* Cocitation of authors results when someone cites any work by any author along with any work by any other author in a new document of his own. It will be seen that in information science such cocitation of authors is highly patterned—to an extent that permits a computer map to be drawn that shows in an intelligible manner the interrelationships of authors in the field. Authors whose works are generally seen to be related, and are repeatedly cited as such in later documents, tend to cluster together on the map, while authors who are rarely or never cited together are relatively far apart. This positioning, it should be emphasized, is based on the composite judgment of hundreds of citers, rather than on any one person's judgment. In effect, it is "the field's view" of 39 authors—or more precisely, the view that emerges from citations in the journal and serial literature (but not in monographs) during the period 1972-1979.

Method

This article combines a series of retrieval tools, developed by White [4], with document mapping tools, devel-

Received May 14, 1980; accepted July 10, 1980

*Thus "Price" in Figure 1 refers not to a man but to a set of writings that includes *Little Science, Big Science; Science Since Babylon;* and many papers. (The use is the same in "I like Hemingway," where "Hemingway" refers to that author's novels, short stories, etc.). The writings of any author that are drawn into the present study through cocitation, as discussed in the Method section, may not be equivalent to his or her total *oeuvre;* for example, Price's writings on early scientific instruments are little represented here because they are not cocited with works by our other 38 choices. They would connect, instead, with works by authors in another field of inquiry, the history of science.

 CCC 0002-8231/81/030163-09$01.80

oped by Small and Griffith [2] and Griffith et al. [3], to map a set of selected authors.

Selection of Authors

The list of 39 authors was taken in considerable part from *Key Papers in Information Science,* edited by Griffith [5], who was attempting to define main areas of the field in terms of recent significant work. The table of contents of that volume supplied 22 names,* and we added 17 more that we judged to be major contributors to the field, including, for perspective, such forerunners as Shannon, Zipf, and Luhn (see Table 1). The resulting list is, of course, biased toward established figures with multiple contributions over the years. It does not purport to capture all of information science, whose contributors number several hundred, but it does include most of the well-known names.

One may want to avoid personal judgment in selecting authors and simply process all cocitation data that are "in the file," recycling from a list of clearly pertinent authors or documents. In that case, one needs special access to Institute for Scientific Information tapes and large amounts of computer money. Alternatively, one may build a citation file of one's own. Our technique, rather, has been personal selection of cited authors and online data gathering through DIALOG. We believe the results are a good first approximation of what we would obtain using more expensive approaches.

The Technique of Cocited Author Retrieval

The logic of cocited author retrieved online has been described in ref. 4. The essence is that in the online citation indexes, Scisearch and Social Scisearch, paired authors' names can be interpreted and entered exactly as if they were ANDed subject headings. Thus, SELECT CR=Crane D? AND CR=Mullins N? means: Select all papers that jointly cite anything by Diana Crane and anything by Nicholas Mullins. ("CR" means "cited reference" and "?" is the truncator whereby one avoids having to specify particular works by Crane or Mullins.) One familiar with these two sociologists' works would expect their paired names to function much like a subject heading, "Sociology of Scientific Communication." And indeed most of the 65 papers (as of December 1979) retrieved on their names would be indexed appropriately under that heading. Note that it is the *combination* Crane-Mullins that has this particular meaning. Since both authors have written on other topics, each of their names alone means something broader.

The advantage of working with cocited authors (as opposed to the cocited documents used in other mappings) is that with nothing more than the authors' names to search on, the cocitation counts for each pair can be obtained online, quickly and at modest cost. The significance of this kind of retrieval as a technical advance is that it permits mapping to order of any small field or specialty an investigator cares to name. One needs only a terminal for online retrieval from Scisearch or Social Scisearch to gather the basic data, and access to widely disseminated computer programs to manipulate them.

To be sure, one can use online search techniques to obtain cocitation counts for specific documents (as opposed to authors), but there are drawbacks to this approach. It requires greater foreknowledge of the field to select the documents–in particular, to select pairs of documents likely to be highly cocited. (Actual cocitation counts are learned only after one gets online.) It requires more bibliographic detail as input to the search, and more time online checking the cited reference (CR=) index with the EXPAND command to search reasonably well. Also, since individual documents are a smaller-scale unit of analysis than authors, as defined here, it is harder to characterize a field or specialty adequately in terms of, say, 30 or 40 documents than in terms of an equivalent number of authors. For example, it is much easier to include "CR= Brookes BC?" as a generic search term, and to mean by that one whole area of information science–all of Brookes' writings and all papers citing them–than it is to pick one or two of his papers that "stand for" the same thing.

TABLE 1. Authors selected for cocited author analysis.

*Allen, Thomas J.	Martyn, John
*Bookstein, Abraham	Meadow, Charles T.
Bradford, S. C.	*Menzel, Herbert
*Brookes, Bertram C.	*Mullins, Nicholas
Cooper, William S.	*Narin, Francis
*Crane, Diana	*Price, Derek de Solla
*Crawford, Susan	*Robertson, Stephen
Fairthorne, Robert A.	*Salton, Gerard
Garfield, Eugene	*Sandison, Alexander
*Garvey, William D.	Saracevic, Tefko
Goffman, William D.	Shannon, Claude E.
*Griffith, Belver C.	*Small, Henry G.
*Jardine, Nicholas	*Sparck Jones, Karen
King, Donald W.	*Swanson, Don
*Kochen, Manfred	*Swets, John
Lancaster, F. Wilfrid	*van Rijsbergen, C. J.
Line, Maurice B.	Vickery, Brian C.
*Lipetz, Ben-Ami	Wilson, Patrick
Luhn, Hans P.	Zipf, George K.
Maron, M. E.	

*Authors appearing in ref. 5.

Data Analysis

Joint postings for all authors pairs were obtained through online intersection of sets.* The resulting data may be thought of as each author's profile of cocitation with every other author on the list. Raw cocitation profiles were keypunched, and the correlation (Pearson's r) of each with

*A brief report on the mapping of these names appears with the frontispiece to the Griffith volume. Some of the volume's authors had not been cited enough overall to have meaningful profiles in a cocitation analysis and so were omitted in that report and here.

*In an effort not to miss postings under varying name forms, Price was entered as "Price D," Desollaprice D," and "Sollaprice D." Brookes was entered both with and without the "e" in his name. All variants yielded hits.

every other was obtained with SPSS. Richter [6] has shown the advantage of normalizing cocitation data through the use of Pearson product-moment correlation coefficients. Briefly, the procedure eliminates scale effects due to the relative citedness of documents, or in this case authors, and instead measures the degree to which cocitations with documents or authors follow a similar profile or pattern. In using a procedure of this type, there was a good deal of difficulty in determining suitable diagonal values to enter into the calculation of the correlation coefficient. In our first analysis, we entered total citations to each author (including self-citations); these scores were sometimes disproportionately large and a few contained homonym errors. (By contrast, the cocitations of *oeuvres*, the author intersections, generate in our tests less than 2% error.) After some thought on the manner in which values in such a matrix might be distributed, we decided that taking the three highest intersections and dividing by two would generate diagonals which would approximate *the next highest score* in the distribution, thus indicating in a general way the relative importance of a particular author within the field.

The next step in the analysis was to enter the Pearson *r*'s between *oeuvres* into nonmetric multidimensional scaling, MDSCAL of Kruskal [7]. The two-dimensional fit was excellent by all normal criteria when the estimated diagonals were used. Orthogonal factor analysis, with rotation to a varimax solution, was also used (see SPSS Manual [8]).

Results

The Map of Authors

Figure 1 gives the MDSCAL map of 39 authors. When they are mapped in two dimensions, as here, stress II of 0.29, coupled with a linear and coherent Shepard plot, indicates a fit of good quality. The two-dimensional plot is clearer when groups of authors are set within boundaries, as shown, by means of Johnson clustering [9]. The map shows:

(1) identifiable author groups (akin to "schools") of information science
(2) locations of these groups with respect to each other
(3) centrality and peripherality of authors within groups and with respect to the overall field
(4) proximities of authors within groups and across group boundaries ("border authors")
(5) positions of authors with respect to the map's axes, which have been set arbitrarily to aid interpretation.

Commentary on each point follows.

Author Groups. From left to right, we see four major groups—one concerned with communications in science and technology; the bibliometricians (mostly British), deriving from Bradford and concerned with statistical properties of subject literatures; four generalists who have provided integrative theory for the field; and the group concerned with problems of automated information retrieval, particularly algorithmic partitioning of document collections. The two

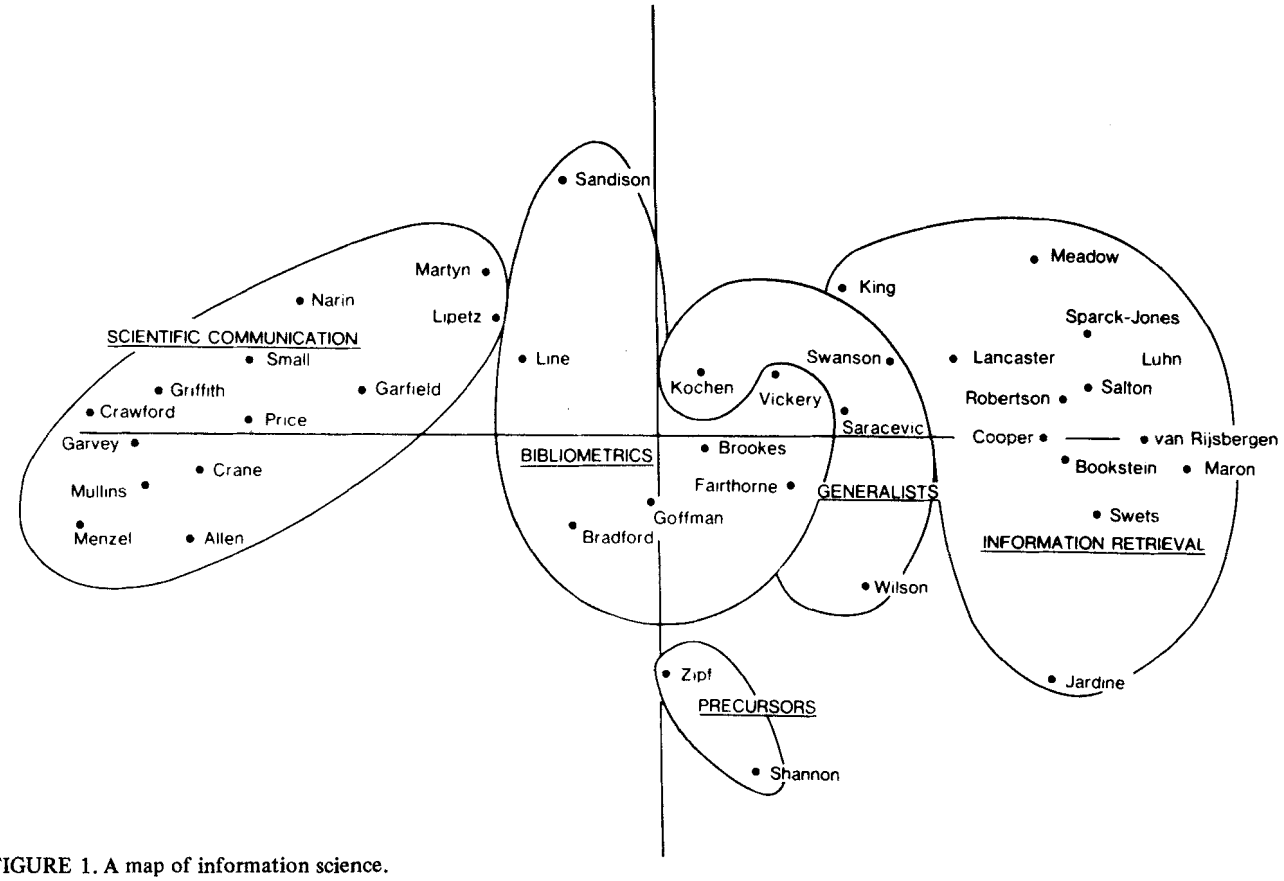

FIGURE 1. A map of information science.

"precursors" from the 1940s are Zipf, whose work on rank-frequency distributions is related to the special case, formulated by Bradford, of the distribution of papers on a topic over journals; and Shannon, provider of a general model of an information system and a measure of "information."

Group Locations. A surprising finding to some may be the centrality of the bibliometrics group. The reason, of course, is that the authors here are cocited with those who write on scientific and technical communications (left) and those who write on information retrieval (right). Those we have labeled generalists have strong links with both the bibliometricians and the retrievalists but not with the authors in sci-tech communications. The latter group and the retrievalists are perceived as having the least in common (in 1972-1979).

Central and Peripheral Authors. The origin of the map is set by the MDSCAL program. Brookes and Kochen lie nearest to it because their works are perceived to share features with many surrounding *oeuvres.* In this sense they are the field's highly representative authors. Within the four main groups on the map, Price's work is central to sci-tech communications, Brookes' to bibliometrics, Saracevic's to general theory, and Salton's to information retrieval, in the sense of being cocited with neighboring *oeuvres* in all directions. Authors near the extremes of the map are related, through cocitation, to fewer neighbors.

Interestingly, the most peripheral author—the one whose cocitation profile yields the fewest high positive correlations with others—is Shannon. As is well known, what he called information theory has been enormously influential in many fields, but cocitation analysis shows that its tie with information science, as of 1980, is tenuous. The works of Luhn and Bradford, among early influences, have more linkages with those of later authors.

Author Proximities. All authors within group boundaries share similar cocitation profiles, reflected in high positive Pearson *r*'s. Closely placed pairs, such as Crane and Mullins in sci-tech communications or Cooper and Bookstein among the retrievalists, have very similar profiles—equivalent to a consensus that the two authors are highly related, usually in subject matter. (However, other relationships, such as similarity of methodology, are possible.) The implication of related pairs for document retrieval is discussed below.

The most distant authors are Crawford and Maron, who have never been cited together and whose cocitation profiles correlate at -0.39. Similar statements hold for other far-distant pairs. However, that two authors are rarely or never cocited, and have negatively correlated profiles, need not imply that citers have consciously judged them dissimilar. It may simply mean that possible relationships between their works (such as between Crawford's and Maron's) have not yet been perceived. This situation could, of course, change over time.

Authors who have high positive correlations with other authors across group boundaries are interesting cases, deserving further study. Both Fairthorne and Vickery, for example, are put with the bibliometricians by the Johnson clustering routine, but they are also highly correlated with the four generalists and with most of the information retrieval group. We have not sorted out the complex reasons that may be present: ties of nationality and common professional forums with other Britishers (Aslib; *Journal of Documentation*); specific contributions to the research tradition begun by Bradford; other contributions aiming at theoretical integration (both have published books, Vickery several); strong emphases in much of their writing on problems of document retrieval. Line presents an analogous case: British, with contributions not only to the Bradford tradition but to studies of the aging of literatures, of library use, of users of information (particularly in the social sciences), and of citation patterns. Thus, his point lies near those of two similarly multifaceted authors in the sci-tech communications group, Lipetz and Martyn.

Oeuvres, we conclude, are inherently ambiguous objects of study. To uncover why they appear on the map as they do, one must make a detailed examination of the writings in them and of how and why they are cocited with writings by other authors. This, of course, involves an analysis at the level of specific documents, which is beyond our present scope, but may be possible in other contexts.

Axes. The vertical axis in Figure 1 seems to represent different variables in the left, center, and right sectors. At the left, lower writers use behavioral measures to study scientific communication; higher writers use citation measures. In the center there is general opposition of theory (Shannon and Zipf) and practice (Sandison and those involved in systems design and evaluation). At the right, those who evaluate retrieval oppose those who develop methods of document analysis.

Left to right on the horizontal axis, there is a transition from authors who use historical time (as in patterns of communication in science) as a variable in their work to authors who do not (as in retrieval), with the shift occurring at about the origin. In other words, works to the left of the origin generally involve time-bound models; works to the right, time-free models.

Works to the left in Figure 1 are "historical" in that they examine relationships among actual researchers, specialties, disciplines, papers, literatures, journals, etc.; the fact that all of these relationships can change over datable periods is present implicitly or explicitly. (It is easy to see why citation indexes, which yield date-bearing clues to changing social and intellectual structures, as reflected in document use, are so important to several of the authors at the left.) Even works in bibliometrics are histories of sorts. In the same way that demographers write very specialized histories of peoples, bibliometricians write specialized histories of subject literatures: their growth, their time-bound distribution over journals, and their rate of aging, as measured by changing citation counts.

To the right of the origin, the authors generally do not "do history" but work with relatively timeless subject matter: e.g., linguistic relationships (particularly between document texts and their representation in assigned or

derived indexing languages); effects of partitioning document collections under various kinds of indexing and various algorithms; discrimination of better effects from worse effects; and the psychology (rather abstract and decision-theoretic) of users of retrieval systems. One finds here a good deal of analysis of concepts: "subject of a document," "relevance of a document to a request," "information need," and so on. There is also an experimental tradition that links performance of retrieval systems with user judgments. Both conceptual analysis and experimentation, however, are deliberately set apart from the study of actual uses of documents by historical persons and in this sense may be said to be "time-free."

Factor Analysis of Authors

Pearson *r*'s for each pair of authors were factor analyzed (SPSS program; principal components with varimax rotation), in an additional search for major interpretable patterns. The correlation data reduced to just seven factors as shown in Table 2. They are listed in order of total variance accounted for, factor 1 being the highest. We have labeled them impressionistically and ranked the authors loading most heavily on each, 0.40 being an arbitrary minimum cutoff point. All 39 authors load on at least one factor; fourteen load on two, and one (Fairthorne) on three, as shown by subscripts to their names.

The factor analysis generally confirms the groupings produced by Johnson clustering (cf. factors 1-4), but also yields some interesting "cross-boundary" groupings not captured by that routine or shown on the map (cf. factors 5-7).

Authors loading on factor 1 are retrievalists, identified in one way or another with automated document analysis. Factor 5 picks out a subset of the retrievalists and shows their linkage with three of the American generalists, plus Fairthorne. This seems to be the group that has been concerned with indexing theory, quantitative evaluation of retrieval, and explication of concepts such as "relevance." Factor 7 also cuts across group boundaries, uniting retrievalists with generalists and with Lipetz. These authors may be perceived as taking a systems-analytic approach to the evaluation of total information systems.

Factor 3 is clearly the bibliometricians—all British except Goffman. (Even he published important work in the British journal *Nature*.) The links between Bradford, his followers, and Zipf are apparent in factor 6, onto which Shannon also loads strongly. Attempts to find general integrative models for information science probably cause the high positive correlation of Shannon and Zipf, who otherwise seem a rather disparate pair.

The remaining factors break out two major configurations in the sci-tech communications group. Factor 2 authors are concerned primarily with networks of papers and journals, particularly citation networks. Factor 4 appears to be simply other behavioral studies of communi-

TABLE 2. Author factor loadings at 0.40 or higher (decimals omitted).[a]

Factor 1 Automated Retrieval, Document Analysis		Factor 2 Networks of Papers; Citation Studies		Factor 3 Bibliometrics		Factor 4 Sci-tech Communications	
van Rijsbergen	89	Small	86	Line	84	Allen	90
Sparck Jones	87	Garfield	84	$Brookes_1$	81	Menzel	87
Salton	86	$Price_1$	78	Sandison	78	Garvey	84
Jardine	75	$Griffith_1$	77	Vickery	70	$Crane_2$	75
$Lancaster_1$	64	Narin	72	$Bradford_1$	62	$Mullins_2$	62
Luhn	59	$Crane_1$	57	Martyn	69	$Crawford_2$	59
$Maron_1$	57	$Mullins_1$	52	$Goffman_1$	50	$Price_2$	54
$Meadow_1$	49	$Crawford_1$	50	$Fairthorne_1$	46	$Griffith_2$	46
$Cooper_1$	43						

Factor 5 Retrieval Evaluation, Indexing		Factor 6 Older Theoretical Models		Factor 7 Evaluation of Information Systems	
$Cooper_2$	78	Zipf	89	$Lancaster_2$	69
Swets	64	Shannon	87	King	62
$Maron_2$	62	$Fairthorne_3$	65	$Swanson_2$	58
Bookstein	60	$Goffman_2$	64	Kochen	58
Robertson	58	$Bradford_2$	56	Lipetz	58
Wilson	57	$Brookes_2$	45	$Saracevic_2$	52
$Saracevic_1$	53			$Meadow_2$	47
$Swanson_1$	49				
$Fairthorne_2$	41				

[a]Subscripts: 1 = first appearance, 2 = second appearance, 3 = third appearance.

cation, developed within the broader framework of the sociology of science and technology. There is considerable overlap between 2 and 4, since several of the authors have contributed to both citation and "other" research within this tradition.

Graphical plots of all possible pairs of rotated factors (a feature of SPSS) show correlation between them as they are grouped in the three preceding paragraphs: first, 1, 5, and 7; second, 3 and 6; third, 2 and 4. The bibliometrics group–factor 3–is correlated to varying degrees with all other factors, as might be inferred from its central position on the MDSCAL map. In contrast, plots of factor 1, the "purest" representation of the retrievalists, with either factor 2 or 4, representing sci-tech communications, show little or no correlation; the loadings are virtually orthogonal. As is seen in Figure 1, the retrievalists and the communications group seem to be perceived as doing different things, as yet unrelated by theory.

Intercorrelations among Authors

The patterns of high positive Pearson *r*'s underlying the factor analysis and the MDSCAL map may be seen in Table 3. The *r*'s have considerable human interest value in revealing who is related to whom, as citers see it, and how strongly. Table 3 also deepens one's understanding of the structure of information science as a field.

Pearson *r*'s close to the diagonal are generally high and fall off away from the diagonal. Those below 0.30 are excluded so that the main patterns emerge. Immediately apparent is the density of intercorrelation among authors in each of the ruled-off areas, corresponding to bounded groups on the map. Apparent, too, is the absence of positive correlations between the sci-tech communications and information retrieval groups (cf. the "great blank" at lower left). A weak link of 0.30 between Lipetz and Lancaster is as close as these groups come to interconnection.

Such links between groups as do appear are attributable in large part to relatively few names: Martyn, Lipetz, and Line; Fairthorne and Vickery; and the generalists Kochen, Swanson, Saracevic, and Wilson. The latter four in particular have interesting correlation patterns: note their many links with both the bibliometrics group (reading their row entries) and the information retrieval group (reading their column entries). The most highly intercorrelated author, simply counting Pearson *r*'s of 0.30 or higher, is Saracevic with 21. (Shannon, as noted earlier, is least intercorrelated with 4.)

Cocitations of Paired Authors

Pearson *r*'s reveal degrees of overall likeness in citation profiles but obscure the actual cocitation counts between pairs of authors. The latter, set forth in Table 4, show information science in a somewhat different light from the matrix in Table 3. Of the 39(38)/2 = 741 possible pairs that can be formed with our authors' names, 479 or 65% have in fact been formed, as of December 1979, through cocitation; i.e., 479 pairs of names will retrieve at least one citing document.

While exact standards of comparison are lacking, 65% connectedness does not seem low; document cocitation matrices of similar size are far less interconnected [10]. Lancaster and Goffman have been cocited at least once with everyone else on the list. Garfield, Brookes, Price, Salton, and Saracevic are each cocited with 37 out of 38 others, and Line, Vickery, Kochen, Narin, and Zipf are only slightly less connected. Several authors' connections are not only wide but "deep" in that they include many instances of double digits. (These are "deep" for information science, but not for many specialties of the natural sciences where cocitation counts in the hundreds are commonplace.) The pair Crane-Price has the highest single count–119. High counts across main group boundaries include Price-Brookes, 29 and Vickery-Lancaster, 25. The dominant impression from Table 4 is that the field is quite unified and coherent but involves only modest levels of citation.

At the same time, hundreds of potential pairs of cocited authors' names in Table 4 suggest nothing more specific than "information science," and so are not very useful as implicit subject indexing for the citing documents they retrieve. (What does "Crane-Luhn" mean? "Shannon-Garvey?" etc.) As a rule we find ourselves driven back to the clusters of high positive Pearson *r*'s (or to the associated factor analysis) for combinations of names that are interpretable as subject indexing at usefully specific levels. Even then, not all combinations are equally suggestive. But taking highly correlated pairs, for example, Allen-Menzel, might be read as "Communication in R&D Settings"; Cooper-Wilson as "Theory of Relevance"; Line-Sandison as "Aging of Literatures"; and Crane-Price as "Invisible Colleges." Such readings may well be confirmed by actual retrievals, in the sense that the latter cite the works by the input pair that one expected and are themselves written in the same general vein.

Outside the highly correlated pairs, it is occasionally possible to place constructions on names linked only by nonzero cocitation counts. For example, the 29 papers retrievable on Price-Brookes, whose correlation is only 0.18, turn out to be largely bibliometric in nature, as might be expected given Brookes' work and that side of Price.

Discussion

Our map of information science suggests some new areas to be explored. Many authors not included on the map (e.g., Buckland or Leimkuhler) can be assigned plausibly to one of its regions, but our assignments have not in fact been tested. Nor is it by any means certain that the Johnson clustering of the 39 authors has identified all the major groupings that exist. There are certain unmapped specialties, such as economics of information, that we think link up with existing groupings (and we can surmise where), but, again, their actual place has not been determined.

Another open question is the degree to which the cocitation structures correspond to social structures. It is our

TABLE 3. Positive Pearson r correlations between authors at 0.30 or higher (decimals omitted).

	Garvey	Menzel	Allen	Crawford	Crane	Mullins	Griffith	Price	Small	Garfield	Narin	Martyn	Lipetz	Line	Sandison	Fairthorne	Vickery	Goffman	Brookes	Bradford	Zipf	Shannon	Kochen	Swanson	Saracevic	Wilson	van Rijsbergen	Jardine	King	Lancaster	Meadow	Swets	Cooper	Robertson	Bookstein	Maron	Salton	Sparck Jones
Menzel	79																																					
Allen	80	94																																				
Crawford	73	54	57																																			
Crane	84	66	66	77																																		
Mullins	71	55	50	68	91																																	
Griffith	65	49	44	73	79	77																																
Price	73	52	54	70	88	74	82																															
Small	42			53	55	51	87	74																														
Garfield	42			45	58	40	65	85	77																													
Narin	32			39	43	33	51	69	65	85																												
Martyn								47	34	64	55																											
Lipetz								41	36	51	40	64																										
Line								45		56	37	77	58																									
Sandison										37		51		80																								
Fairthorne													33	44	30																							
Vickery												48	34	61	47	78																						
Goffman														37	33	76	63																					
Brookes												46		63	63	80	82	86																				
Bradford												38		45	42	71	72	86	90																			
Zipf																58	33	47	38	45																		
Shannon																52		46		34	92																	
Kochen												39	58	46		69	58	51	53	46	32																	
Swanson													36			39	37						51															
Saracevic													39	34		66	61	61	54	43			74	68														
Wilson																39	31							38	47													
van Rijsbergen																								34														
Jardine																											81											
King																42							43	57	52		37											
Lancaster													30			33	50						49	73	66		59	34	81									
Meadow																								48	35		50		59	71								
Swets																								45	42		50		30	43								
Cooper																37	30						35	65	65	64	62		55	67	50	71						
Robertson																35	38		31					50	52	37	71	36	47	68	38	81	84					
Bookstein																45	39						38	60	52	39	57		47	66	52	76	74	74				
Maron																								47	38	36	67	38		51	46	72	77	75	81			
Salton																	31						32	63	55		82	59	62	85	68	58	71	77	73	71		
Sparck Jones																								39	35		80	64	40	73	51	40	49	66	51	57	87	
Luhn																								40	32		50	32		57	51	41	39	45	54	66	72	68

Scientific Communications

Bibliometrics

Precursors

Generalists

Information Retrieval

TABLE 4. Cocitation counts between author pairs.

	Garvey	Menzel	Allen	Crawford	Crane	Mullins	Griffith	Price	Small	Garfield	Narin	Martyn	Lipetz	Line	Sandison	Fairthorne	Vickery	Goffman	Brookes	Bradford	Zipf	Shannon	Kochen	Swanson	Saracevic	Wilson	van Rijsbergen	Jardine	King	Lancaster	Meadow	Swets	Cooper	Robertson	Bookstein	Maron	Salton	Sparck Jones
Menzel	20																																					
Allen	23	31																																				
Crawford	11	4	8																																			
Crane	37	23	27	18																																		
Mullins	19	11	8	9	65																																	
Griffith	18	13	9	14	42	32																																
Price	49	24	29	21	119	49	63																															
Small	8	2	3	7	19	16	46	49																														
Garfield	7	2	4	6	26	10	30	118	46																													
Narin	6	1	4	5	11	7	10	30	15	42																												
Martyn	8	4	5	0	5	1	3	17	6	24	9																											
Lipetz	4	2	5	0	4	1	5	9	5	9	4	8																										
Line	8	6	9	2	11	2	4	31	6	23	7	18	8																									
Sandison	0	0	0	0	0	0	0	8	0	15	2	5	1	20																								
Fairthorne	0	0	1	0	3	1	1	3	1	7	4	2	6	11	2																							
Vickery	3	0	4	0	1	1	3	17	4	9	1	12	4	18	8	16																						
Goffman	3	2	1	3	10	5	5	25	6	17	5	4	3	7	4	16	16																					
Brookes	2	0	3	1	6	2	3	29	2	23	7	11	5	30	20	24	35	41																				
Bradford	3	0	0	1	4	1	4	24	5	23	5	8	2	6	3	15	22	28	47																			
Zipf	0	1	3	1	2	2	3	14	2	6	1	3	3	3	0	11	8	10	10	12																		
Shannon	1	1	1	0	1	0	1	7	1	3	2	0	0	2	0	11	2	13	9	6	27																	
Kochen	2	1	0	1	5	1	1	13	2	12	3	3	6	7	0	10	8	8	11	6	4	4																
Swanson	2	3	2	0	4	1	0	9	0	4	1	0	6	3	1	6	6	4	9	1	0	1	5															
Saracevic	3	4	1	0	5	1	2	15	4	11	3	5	6	11	1	11	12	19	15	12	4	6	11	11														
Wilson	3	1	3	1	1	1	2	3	2	2	1	2	0	5	0	3	4	3	3	4	1	1	2	3	5													
van Rijsbergen	1	0	0	0	0	0	1	2	1	1	0	0	0	1	0	1	3	5	5	0	2	0	1	2	4	1												
Jardine	0	0	0	1	1	0	3	3	3	2	1	0	0	0	0	0	0	3	2	0	2	3	0	1	1	1	14											
King	5	0	5	1	2	2	3	10	1	7	3	4	2	4	0	3	10	2	5	2	3	0	5	5	8	1	4	3										
Lancaster	6	1	9	3	4	2	4	16	5	15	3	16	8	15	1	14	25	8	16	4	7	1	14	20	23	6	13	3	27									
Meadow	0	0	1	0	1	0	1	1	0	1	1	1	0	2	0	1	3	1	2	0	1	2	1	2	2	0	4	0	4	14								
Swets	0	1	0	0	0	0	0	3	0	1	0	1	1	2	0	8	4	9	17	2	2	6	1	5	5	2	5	2	5	9	0							
Cooper	0	2	4	0	3	0	1	7	0	3	1	2	1	6	0	9	8	11	10	2	2	3	6	10	15	10	11	2	10	20	5	11						
Robertson	1	0	1	0	0	0	0	5	1	4	3	10	0	8	2	7	8	10	18	2	1	2	3	5	9	3	13	3	7	20	0	15	20					
Bookstein	0	0	2	0	2	1	2	7	1	2	3	2	5	9	0	9	8	6	14	4	6	1	6	7	8	4	8	2	6	18	4	12	12	12				
Maron	0	0	0	0	0	0	1	2	0	0	0	0	1	2	0	4	4	7	6	1	2	1	5	7	8	5	9	3	3	8	5	12	16	15	17			
Salton	1	1	4	0	1	2	4	11	6	14	3	3	5	5	1	10	14	14	18	5	13	9	8	16	20	3	26	20	15	60	10	15	21	28	22	27		
Sparck Jones	1	0	0	0	0	0	1	3	3	6	2	3	2	3	1	2	10	3	5	1	2	0	4	3	6	1	14	9	3	21	3	5	7	13	6	8	36	
Luhn	1	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	1	2	3	0	1	2	1	2	4	0	2	0	0	10	2	3	2	2	5	9	16	8

impression that authors grouped together by the clustering routine may, as persons, have more ties of communication within their groups than outside them. This seems particularly true of the two most distant groups—sci-tech communications and the retrievalists. One writer, upon seeing the map, said he knew everyone about two-thirds across the map and then no one.

In conclusion, we believe that we have presented an interesting trial of a technique that might contribute much to the understanding of intellectual structure and movement in the sciences and in other areas, to the extent that they rely on serial publication. We should add that we can group citing documents on the basis of age, using ISI acquisition numbers, and thus examine change over time. This tool seems to establish authors, in addition to documents, as an interesting unit in the analysis of ideas.

Acknowledgments

The authors thank Lynn A. Deglin for many forms of assistance in tabulating and displaying data and creating the manuscript. Griffith, Deglin, and the costs of data collection and computation were supported by Public Health Service Grant from the National Library of Medicine, No. R01 LM00911-07.

References

1. Small, H. G. "Co-citation in the Scientific Literature: A Measure of the Relationship between Two Documents." *Journal of the American Society for Information Science.* 24:265–269; 1973.
2. Small, H. G.; Griffith, B. C. "The Structure of Scientific. Literatures I: Identifying and Graphing Specialties." *Science Studies.* 4:17–40; 1974.
3. Griffith, B. C.; Small, H. G.; Stonehill, J. A.; Dey, S. "The Structure of Scientific Literatures II: Toward a Macro- and Microstructure for Science." *Science Studies.* 4:339–364; 1974.
4. White, H. D. "Cocited Author Retrieval Online: An Experiment with the Social Indicators Literature." *Journal of the American Society for Information Science.* 32:16–22; 1981.
5. Griffith, B. C., Ed. *Key Papers in Information Science.* White Plains, NY: Knowledge Industry Publications; 1980.
6. Richter, S. B. Personal communication, 1979.
7. Kruskal, J. B. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika.* 29:1–27; 1964; Kruskal, J. B. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika.* 29:115–129; 1964.
8. Nie, N. H.; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K.; Bent, D. H. *SPSS: Statistical Package for the Social Sciences, 2nd ed.* New York: McGraw-Hill; 1975.
9. Johnson, S. "Hierarchical Clustering Schemes." *Psychometrika.* 32:241–254; 1967. (For copies of the dendrogram of all 39 names, contact H. White.)
10. Small, H. G.; Griffith, B. C. Personal communication, 1980.