Journal of the American Society for Information Science and Technology, Volume 54 (5): 400--412 , March 2003

Why Do We Need Algorithmic Historiography?

Eugene Garfield
Institute for Scientific Information, 3501 Market Street, Philadelphia, PA 19104.
E-mail: Garfield@codex.cis.upenn.edu

A. I. Pudovkin
Institute of Marine Biology, Russian Academy of Sciences, Vladivostok 690041, Russia.
E-mail: aipud@online.ru

V. S. Istomin
Center for Teaching, Learning, and Technology, Washington State University, Pullman, WA 99164-4550.
E-mail: Vi@mail.wsu.edu

This article discusses the rationale for creating historiographs of scholarly topics using a new program called *HistCite*™, which produces a variety of analyses to aid the historian identify key events (papers), people (authors), and journals in a field. By creating a genealogic profile of the evolution, the program aids the scholar in evaluating the paradigm involved.

Introduction

The editor of this special issue, Chaomei Chen, asked us to provide an applications-oriented view of algorithmic historiography and how it relates to understanding scientific paradigms. The first public presentations of our system for algorithmic historiography were given at the University of Pittsburgh (Garfield, 2001a) and then at Drexel University (Garfield, 2001b). In addition, a paper by the authors has been presented at the national meeting of the American Society for Information Science and Technology (Garfield, Pudovkin, & Istomin, 2002).

Of late, it has become fashionable to speak about paradigms, but a paradigm is basically a model of a field. Historiography is intimately related to understanding paradigms shifts. However, we tend to use the term for concepts that are broader in scale than typical literature searches. But if you want to understand how a paradigm has changed, you must identify the literature of the topic first. Then by observing the changes in the citation of key works of that field, you see how the basic concepts or the perception of the paradigm changed.

Our historiographic software, hereinafter referred to as *HistCite*™, facilitates the understanding of paradigms by enabling the scholar to identify the significant works on a given topic. At the same time it provides a graphic, geneologic presentation of the citational links between them. The system also preserves these citation links for the entire bibliographic collection so that one can explore intermediary links involving less frequently cited works.

The historiography of science can be viewed in a variety of ways. Classical historians of science discuss ancient, medieval, renaissance, and contemporary 19th and 20th century science descriptively. In their narratives and analysis, they also provide extensive bibliographic documentation. Historic scholars are frequently heavy users of library services and usually take great pride in having read both the classical and ancillary, even ephemeral works in the field. In our conception of facilitating historiography—that is, writing the history of modern science—we make the basic assumption that the bibliographic information contained in a collection of published scientific articles is sufficient for the purpose of recapturing the historiographic structure of the field. Because citation indexes utilize the cited works of thousands of authors, it is assumed that collectively they call out the basic works in any field. Gaps may exist in the documentation provided by individual authors, but collectively they produce a fairly complete picture of the historic background of the topic.

Our approach to historiography is decidedly genealogic. We want to show where a particular topic began and identify both the bibliographic antecedents and descendents of its principal, often primordial papers and authors. Once these basic structural elements (papers and books) of the field are identified, they are "summarized" graphically as an

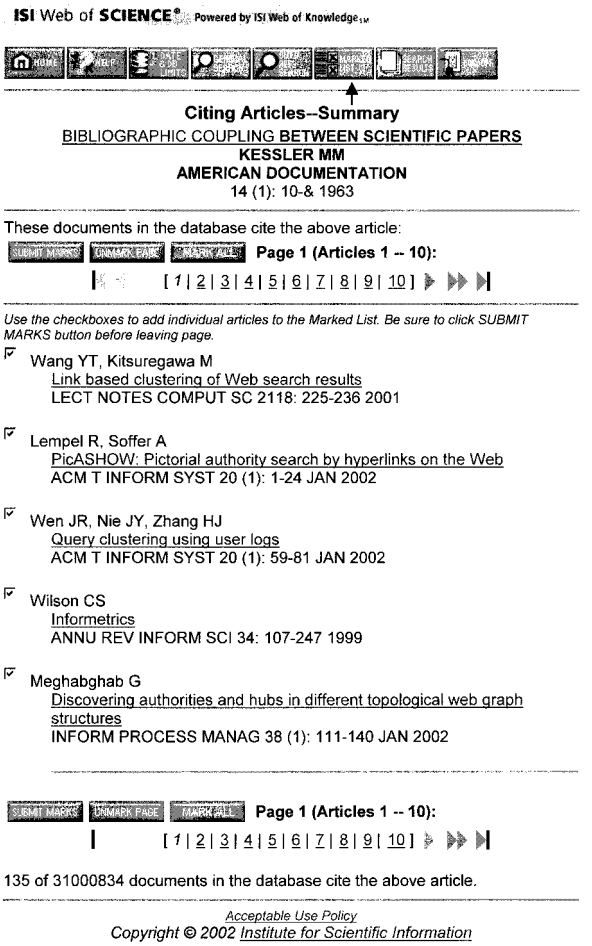


FIG. 1. *Web of Science* cited reference search on MM Kessler 1963.

interconnected historiograph involving, typically, the 5% that are the most-cited. A chronologic map or flow diagram is created in which the nodes for these works are represented either as circles, ovals, rectangles, or other icons. The

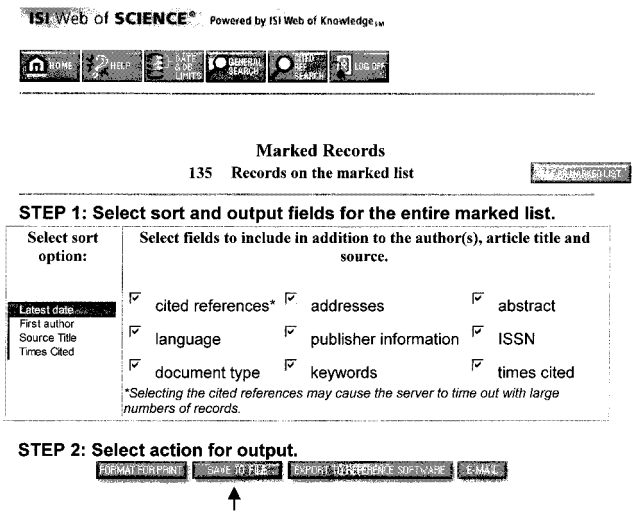


FIG. 2. Marked records field options and save to file.

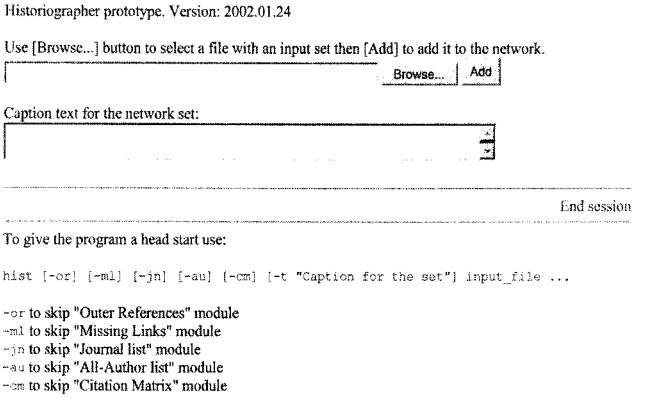


FIG. 3. Start screen. File address and caption are inserted and Add pressed to activate *HistCite* software.

first step in constructing this map is to compile a bibliography from the *ISI Web of Science*® or other database based on a search by keywords, cited authors, cited papers or books, or by a source journal. In a typical search, you obtain a list of citing papers (Fig. 1). You click on Marked List (XX) and the export file is saved in ASCII format (Fig. 2). When the *HistCite* software is opened up, a screen appears

Outer References Missing Links? Journal list All-Author list Citation Matrix

Citations to Kessler's Bibliographic Coupling and papers with BC in title/abstract
Nodes: 223
Sorted by **year, journal, volume, page**.

Cited nodes	Nodes / Authors	GCS	LCS
0	1 1963 AMERICAN DOCUMENTATION 14(1):10-& **KESSLER MM** *Bibliographic Coupling Between Scientific Papers*	128	134
1	2 1963 AMERICAN DOCUMENTATION 14(4):289-& **GARFIELD E** *Citation Indexes in Sociological and Historical Research*	61	6
0	3 1963 IEEE TRANSACTIONS ON INFORMATION THEORY 9(1):49-& **KESSLER MM** *An Experimental Study of Bibliographic Coupling Between Technical Papers*	8	9
1	4 1963 INFORMATION STORAGE AND RETRIEVAL 1(4):169-187 **KESSLER MM** *Bibliographic Coupling Extended in Time-10 Case Histories*	14	15
1	5 1964 ASLIB PROCEEDINGS 16(2):48-63 **[Anon]** *Aslib 37th Annual Conference - University of St Andrews, 24th- 26th September 1963*		0
1	6 1964 ASLIB PROCEEDINGS 16(4):132-152 **LANCASTER FW** *Mechanized Document Control - A Review of Some Recent Research*	3	0
2	7 1964 ASLIB PROCEEDINGS 16(8):246-251 **KEEN EM** *Citation Indexes*	5	2
0	8 1964 JOURNAL OF DOCUMENTATION 20(4):236-236 **MARTYN J** *Bibliographic Coupling*	12	7
1	9 1964 NACHRICHTEN FUR DOKUMENTATION 15(3):122-130 **MODEL F** *Citation Index and Retrospective Cataloging - Examples of Citation Documentation*	5	0
1	10 1964 SCIENCE 144(361):649-& **GARFIELD E** *Science Citation Index-New Dimension in Indexing - Unique Approach Underlies Versatile Bibliographic Systems for Communicating + Evaluating Information*	92	19
3	11 1965 AMERICAN DOCUMENTATION 16(3):223-233 **KESSLER MM** *Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing*	42	43

FIG. 4. Chronologic output file after *HistCite*™ processing.

MARSHAKOVA IV

78 1973 NAUCHNO-TEKHNICHESKAYA INFORMATSIYA SERIYA 2-INFORMATSIONNYE PROTSESSY I SISTEMY 2(6):3-8
MARSHAKOVA IV
System Of Document Connections Based On References

113 1981 SCIENTOMETRICS 3(1):13-25
MARSHAKOVA IV
Citation Networks In Information-Science

118 1982 NAUCHNO-TEKHNICHESKAYA INFORMATSIYA SERIYA 2-INFORMATSIONNYE PROTSESSY I SISTEMY (2):1-5
MARSHAKOVA IV
Determination of Tendencies in the Development of Science and Technology by the Analysis of Documentary Data

FIG. 5. Papers published by I.V. Marshakova including 1973 primordial paper on cocitation clustering.

that asks you to add the saved file address. You click the browse button and select the appropriate file address and then click add (see Fig. 3). You can also add a caption text for the network. The address for the saved file is used to activate the *HistCite* software. Once the program is activated, processing is completed in about one minute for a file of 500–1,000 source records. It is essential to reiterate that for each source record captured, its entire list of cited references must be included. Although records without cited references will be included in the final bibliography, they could not by definition provide backward links.

The *Science Citation Index® (SCI®), Social Sciences Citation Index® (SSCI®)*, and *Arts and Humanities Citation Index® (AH&CI®)* on CD ROM, on the *Web of Science*, or on Dialog can be used to create the import file. Once completed, the program displays the output in the form of a chronological table arranged in order by year, journal, volume, and page. A typical display is shown in Figure 4. Each item is assigned a node serial number, and for each paper there is a *Global Citation Score* (*GCS*) as well as a *Local Citation Score* (*LCS*). By clicking on either *GCS* or *LCS*, the researcher is presented with a ranked sort of the collection starting with the most-cited paper at the top. The *GCS* is the number of times the paper is cited in the *SCI*. The *LCS* is the number of times the paper is cited in the collection.

ISI Web of Science location:
Cited references outside of this network.
Total: 6320 (top 100 shown).

LCS	Reference
31	PRICE DJD, 1965, SCIENCE, V149, P510 WoS
25	GARFIELD E, 1955, SCIENCE, V122, P108 WoS
17	GARFIELD E, 1979, CITATION INDEXING WoS
17	GARFIELD E, 1964, USE CITATION DATA WR WoS
16	GRIFFITH BC, 1974, SCI STUD, V4, P339 WoS
14	GARFIELD E, 1972, SCIENCE, V178, P471 WoS
14	SALTON G, 1983, INTRO MODERN INFORMA WoS
14	MARGOLIS J, 1967, SCIENCE, V155, P1213 WoS
14	MORAVCSIK MJ, 1975, SOC STUD SCI, V5, P86 WoS
13	CRANE D, 1972, INVISIBLE COLLEGES WoS
13	SALTON G, 1963, J ACM, V10, P440 WoS
12	SMALL HG, 1977, SOC STUD SCI, V7, P139 WoS
11	NARIN F, 1976, EVALUATIVE BIBLIOMET WoS
10	WHITE HD, 1981, J AM SOC INFORM SCI, V32, P163 WoS
10	KUHN TS, 1970, STRUCTURE SCI REVOLU WoS
10	WESTBROOK JH, 1960, SCIENCE, V132, P1229 WoS
10	PRITCHARD A, 1969, J DOC, V25, P348 WoS
10	KWOK KL, 1975, INFORMATION PROCESSI, V11, P201 WoS
10	KAPLAN N, 1965, AM DOC, V16, P179 WoS
10	CHUBIN DE, 1975, SOC STUD SCI, V5, P423 WoS
10	LOTKA AJ, 1926, J WASHINGTON ACADEMY, V16, P317 WoS
9	GARFIELD E, 1979, CITATION INDEXING IT WoS
9	VANRIJSBERGEN CJ, 1979, INFORMATION RETRIEVA WoS
9	SMALL HG, 1978, SOC STUD SCI, V8, P327 WoS
9	NARIN F, 1972, J AM SOC INFORM SCI, V23, P323 WoS
9	GARFIELD E, 1980, CURRENT CONTENT 0609, P5 WoS
9	PRICE DJD, 1966, AM PSYCHOL, V21, P1011 WoS
8	GARFIELD E, 1970, NATURE, V227, P669 WoS
8	BUSH V, 1945, ATLANTIC MONTHLY, V176, P101 WoS
8	PINSKI G, 1976, INFORMATION PROCESSI, V12, P297 WoS

FIG. 6. Outer references ranked by citation frequency.

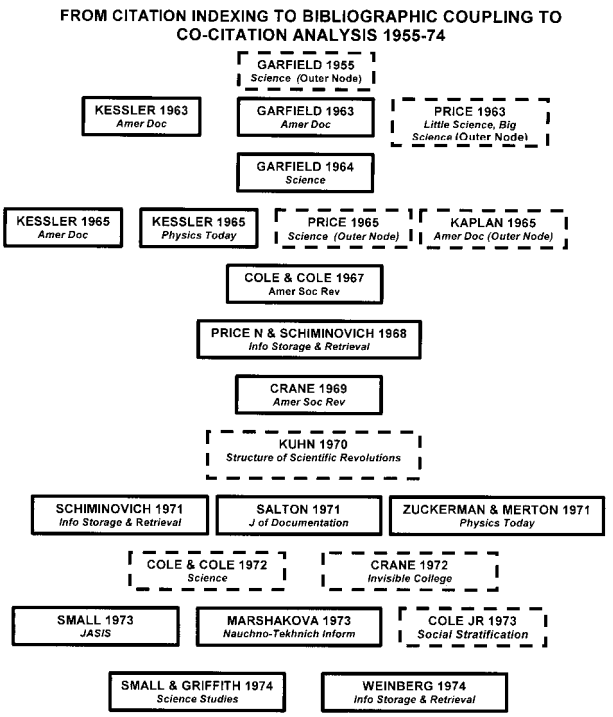


FIG. 7. Historigraph of bibliographic coupling from 1955 to 1974. Dotted lines indicate outer references.

Although there are exceptions, it has been well documented that the key works in most fields are well correlated with citation frequency. To construct the history of bibliographic coupling Garfield (Garfield, 2001a) began the *WoS* cited reference search with the primordial paper by Michael M. Kessler (Kessler, 1963). That paper can be regarded as a symbol for the topic of bibliographic coupling. Nevertheless, for whatever reasons, some authors either have forgotten the paper or do not think it necessary to cite Kessler's paper even though the titles of their papers include the term bibliographic coupling. So the search was augmented with a title-word search that added a few dozen papers. The final bibliography contains a few hundred source papers, each of which has cited an average of 20 or more references. Thus, a virtual minicitation index consisting of 4–5,000 cited references was created.

Many citation variations are due to lack of standardization in citation practice. To capture all relevant papers one may have to modify the *WoS* search profile accordingly. For

Outer References Missing Links? Journal list All-Author list Citation Matrix

Geneflow Papers - 1974 to August 2001
See the Historiograph of the 29 most cited papers in LCS by clicking here
Nodes: 620
Sorted by **year, journal, volume, page**.

Cited nodes	Nodes / Authors	GCS	LCS
0	1 1974 GENETICS 78(3):961-965 **SPIETH PT** *Gene Flow and Genetic Differentiation*	43	9
0	2 1975 AMERICAN NATURALIST 109(969):597-601 **SLATKIN M; MARUYAMA T** *Influence of Gene Flow on Genetic Distance*	21	6
0	3 1975 AMERICAN NATURALIST 109(970):659-676 **MAY RM; ENDLER JA; MCMURTRIE RE** *Gene Frequency Clines in Presence of Selection Opposed by Gene Flow*	88	15
0	4 1975 AUK 92(3):493-510 **COOKE F; MACINNES CD; PREVETT JP** *Gene Flow Between Breeding Populations f Lesser Snow Geese*	71	3
0	5 1975 GENETICS 80(2):349-361 **MCKENZIE JA** *Gene Flow and Selection in a Natural Population of Drosophila- Melanogaster*	17	0
0	6 1975 GENETICS 81(4):787-802 **SLATKIN M** *Gene Flow and Selection in a 2-Locus System*	52	3
0	7 1975 HEREDITY 34(JUN):407-415 **BRUSSARD PF; VAWTER AT** *Population Structure, Gene Flow and Natural Selection in Populations of Euphydryas-Phaeton*	9	0
0	8 1975 JOURNAL OF MOLECULAR EVOLUTION 5(3):177-185 **ADAMS RP** *Gene Flow Versus Selection Pressure and Ancestral Differentiation in Composition of Species - Analysis of Populational Variation of Juniperus-Ashei Buch Using Terpenoid Data*	21	0
0	9 1976 GENETICS 83(3):S64-S64 **RODELL CF** *Some Demographic Considerations of Gene Flow*		0
0	10 1976 JOURNAL OF BIOSOCIAL SCIENCE 8(4):309-333 **DYER KF** *Patterns of Gene Flow Between Negroes and Whites in United- States*	4	0

FIG. 8. Chronologic table of papers on gene flow from 1974 to 2001.

example, some papers do not cite either the first or second initial of the author, in this case M.M. Kessler. So in conducting the *WoS* search, one should anticipate such variants and, if necessary, iterate the search so that as many relevant papers as possible are pulled into the collection. The existence of such variants may not be noted until the first iteration of *HistCite* is finished. *HistCite* also makes it possible to identify some variants by assembling them either in the "missing links" or "outer references" lists. Another variant may be found in changed names. Some authors add or drop initials to their by-lines, as was the case for Henry Small. Married names are also an occasional variant.

Examining these lists will identify candidate papers that have cited a relevant work but do not initially get pulled into the network. A large number of these variants involve cites to specific pages of individual papers or books. Therefore, to be included in the network, the first page of these cited works has to be added, so that they can be matched against the fully paginated record for the source article involved. Initially, during the first iteration of the program, many variations may be observed. Some prove to be redundant because several different pages of the same work may be cited in a single source. Once the pagination is corrected, the citing source is accounted for in the network.

Other variations occur in journal title abbreviations. A particularly egregious example involved citations to a Russian article by Irina Marshakova. Her primordial paper on cocitation analysis is node #78, not shown in Figure 4. It is shown in Figure 5 as part of her author output. This paper was published in *Nauchno-Tekhnicheskaya Informatsiya.* The abbreviation for the Russian journal has been cited variously, as, for example, *NTI*, *Nauch Tekh Inform,* with or without the series number as shown in Figure 5. The full journal title includes a series number, because the journal was published in two sections. Most Western journals published in sections use alphabetic characters. For example, the well-known *Physical Review* is published in six sections: A, B, C, D, E, and F. After editing all the variant citations to Marshakova's paper, the citation count increased from 3 to 20. Fortunately, we cannot demonstrate all of these variations now because they have since been corrected in the *WoS* database.

Although the task of editing and correcting variations is important, it is secondary to the task of examining and editing the list of "outer references" (Fig. 6). These are cited references that do not match any items in the main bibliography. These are sorted by citation frequency. These outer references normally outnumber the nodal bibliography papers by a factor of ten or more because each source paper cites 20 or more works. Large numbers of variations turn up in the outer references. The variants in the file of outer

Outer References Missing Links? Journal list All-Author list Citation Matrix

Gene Flow Papers – 1974 to August 2001
See the Historiograph of the 29 most cited papers in LCS by clicking here
Nodes: 620
Sorted by **LCS**.

Cited nodes	Nodes / Authors	GCS	LCS
11	71 1985 ANNUAL REVIEW OF ECOLOGY AND SYSTEMATICS 16():393-430 **SLATKIN M** *Gene Flow in Natural Populations*	554	111
4	121 1987 SCIENCE 236(4803):787-792 **SLATKIN M** *Gene Flow and the Geographic Structure of Natural Populations*	646	104
2	76 1985 EVOLUTION 39(1):53-65 **SLATKIN M** *Rare Alleles as Indicators of Gene Flow*	536	100
4	153 1989 EVOLUTION 43(7):1349-1368 **SLATKIN M; BARTON NH** *A Comparison of 3 Indirect Methods for Estimating Average Levels of Gene Flow*	401	82
0	37 1981 GENETICS 99(2):323-335 **SLATKIN M** *Estimating Levels of Gene Flow in Natural Populations*	220	53
0	29 1980 NATURE 284(5755):450-451 **SCHAAL BA** *Measurement of Gene Flow in Lupinus-Texensis*	165	39
4	31 1981 ANNALS OF THE MISSOURI BOTANICAL GARDEN 68(2):233-253 **LEVIN DA** *Dispersal Versus Gene Flow in Plants*	190	37
5	112 1987 EVOLUTION 41(2):385-400 **WAPLES RS** *A Multispecies Approach to the Analysis of Gene Flow in Marine Shore Fishes*	198	30
4	64 1984 GENETICS 106(2):293-308 **LARSON A; WAKE DB; YANEV KP** *Measuring Gene Flow Among Populations Having High-Levels of Genetic Fragmentation*	119	28
0	16 1977 THEORETICAL POPULATION BIOLOGY 12(3):253-262 **SLATKIN M** *gene flow and genetic drift in a species subject to frequent local extinctions*	162	25
3	156 1989 GENETICS 123(3):603-613 **SLATKIN M; MADDISON WP** *A Cladistic Measure of Gene Flow Inferred from the Phylogenies of Alleles*	171	22

FIG. 9. Gene flow collection sorted by local citation score (LCS).

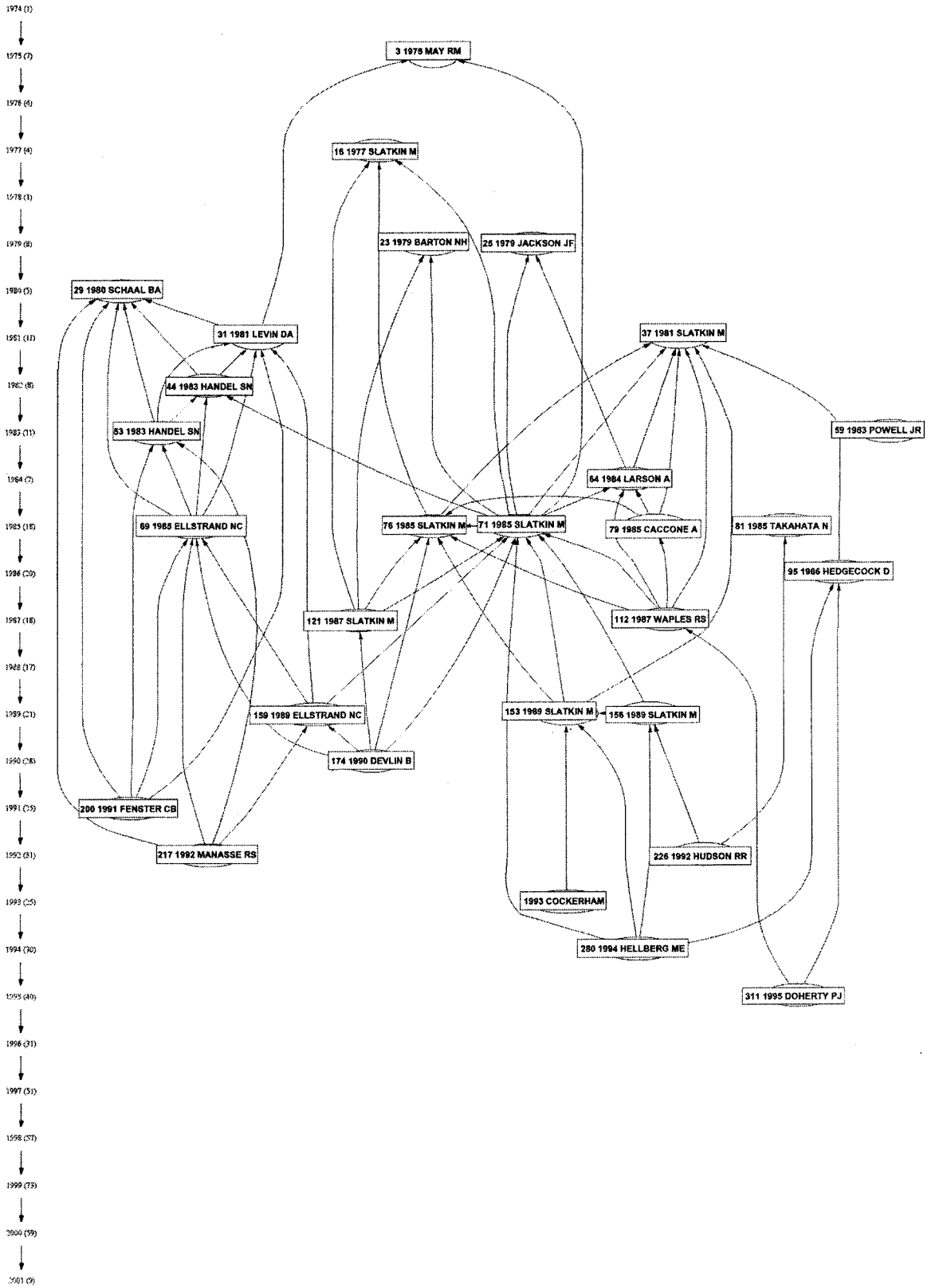


FIG. 10. Graphic presentation of gene flow collection.

references are best identified and edited by sorting them by cited author or by using the ”Find” command of your browser to find all entries for one author at a time. However, of greater significance, the outer references collection provides a potential pool of core papers and books. By definition, these works were not retrieved in the original *WoS*

3

Author(s)	MAY RM; ENDLER JA; MCMURTRIE RE
Title	GENE FREQUENCY CLINES IN PRESENCE OF SELECTION OPPOSED BY GENE FLOW
Journal	AMERICAN NATURALIST 109(970):659-676
Year	1975
Type	Article
Address	PRINCETON UNIV,BIOL DEPT,PRINCETON,NJ 08540
Abstract	
WoS CS	88
LCS	15
cites	0
CR[17]	BARBER HN, 1965, HEREDITY, V20, P551 BARBER HN, 1957, NATURE, V179, P1267 BISHOP J, 1972, J ANIM ECOL, V4, P209 CROW JF, 1970, INTRO POPULATION GEN ENDLER JA, 1973, SCIENCE, V179, P243 ENDLER JA, 1976, SUBSPECIES SPECIES C FISHER RA, 1937, ANN EUGEN, V7, P355 FISHER RA, 1950, BIOMETRICS, V6, P353 HALDANE JBS, 1948, J GENET, V48, P277 HANSON WD, 1966, BIOMETRICS, V22, P453 JAIN SK, 1966, HEREDITY, V21, P407 KETTLEWELL HBD, 1969, HEREDITY, V24, P1 KETTLEWELL HBD, 1969, HEREDITY, V24, P15 KETTLEWELL HBD, 1961, HEREDITY, V16, P403 KIMURA M, 1958, 9 NAT I GEN ANN REP, P84 ROUGHGARDEN J, 1974, AM NAT, V108, P649 SLATKIN M, 1973, GENETICS, V75, P733

FIG. 11. Full source entry for none #3, paper by R.M. May, *American Naturalist*, 1975.

16

Author(s)	SLATKIN M
Title	GENE FLOW AND GENETIC DRIFT IN A SPECIES SUBJECT TO FREQUENT LOCAL EXTINCTIONS
Journal	THEORETICAL POPULATION BIOLOGY 12(3):253-262
Year	1977
Type	Article
Address	UNIV CHICAGO,DEPT BIOPHYS & THEORET BIOL,CHICAGO,IL 60637
Abstract	
WoS CS	162
LCS	26
cites	0
CR[11]	COHEN D, 1976, THEORETICAL POPULATI, V10, P276 CROW JF, 1970, INTRO POPULATION GEN CROWELL KL, 1973, AM NAT, V107, P535 KIMURA M, 1964, GENETICS, V49, P725 LATTER BDH, 1973, GENETICS, V73, P147 LEVIN BR, 1974, EVOLUTION, V28, P527 LEVINS R, 1970, SOME MATH QUESTIONS, V2, P75 MARUYAMA T, 1970, THEOR POPUL BIOL, V1, P273 SIMBERLOFF DS, 1969, ECOLOGY, V50, P278 SLATKIN M, UNPUBLISHED WRIGHT S, 1931, GENETICS, V16, P97

FIG. 12. Full source entry for M. Slatkin's 1977 paper. *WoS* CS is Global Citation Source.

search because they did not cite the primordial work of Kessler. However, they were frequently cited by the papers that did cite Kessler. Indeed, many of these outer references are key works that contributed to the evolving history of the topic. Hence, works by Garfield and others published prior to 1963 will turn up as outer references. However, many papers published after 1963, which have been cocited with Kessler, also prove to be essential to the subject. Indeed, an important by-product of the algorithm is the ability to identify candidate papers for performing a cocitation mapping exercise.

The ultimate hand-drawn map of the topic of bibliographic coupling for 1955–1974 is shown in Figure 7. As can be seen, the map is arranged chronologically. From its citation frequency in this collection, it was determined that Garfield's 1955 paper in *Science* was an essential predecessor to the work on bibliographic coupling even though in his 1963 paper Kessler himself did not cite it. The Garfield paper appeared in the list of *outer references* (Fig. 6), as did the 1963 work by Derek de Solla Price, *Little Science, Big Science.* The book did not explicitly cite the Kessler paper. Upon further investigation, however, it was observed that Price acknowledged Kessler's technical reports in a footnote. By virtue of its inclusion in outer references it was realized that the Price work was essential to a more detailed history of this topic. Historiography involves making judgments and selections. But the basis for these selections must be a comprehensive display of candidate works, and the system also permits one to identify key linking or transition papers that may or may not be well cited.

Even a scholar with a photographic memory like Isaac Asimov can overlook some transitional key works that were involved in the development of a field (Garfield, Sher, & Torpie, 1964).

In a similar fashion, the history of the topic of cocitation analysis was reconstructed (Garfield, 2001b). It was clearly shown that this topic is a logical extension of bibliographic coupling. However, the literature had grown from a few hundred papers to over 1,000. In the evaluation of the topic, the key role of later researchers like Howard White and Katherine McCain could be seen. Indeed, the process goes to the heart of understanding the paradigmatic shift from bibliographic coupling to cocitation to coword analysis.

Outer nodes? Missing links? Journal list Author list

Chronological File for *Evolution*

Nodes: 5112
Sorted by **year, journal, volume, page.**

Cited nodes	Nodes / Authors	GCS	LCS
0	0 1947 EVOLUTION 1(1-2):1-16 **DOBZHANSKY T** *Adaptive Changes Induced by Natural Selection in Wild Populations of Drosophila*	82	21
0	1 1947 EVOLUTION 1(1-2):17-31 **SPIETH HT** *Sexual Behavior and Isolation in Drosophila .1. The Mating Behavior of Species of the Willistoni Group*	20	7
0	2 1947 EVOLUTION 1(1-2):32-41 **STIRTON RA** *Observations on Evolutionary Rates in Hypsodonty*	20	1
0	3 1947 EVOLUTION 1(1-2):42-47 **IVES PT** *2nd Chromosome Inversions in Wild Populations of Drosophila- Melanogaster*	22	5
0	4 1947 EVOLUTION 1(1-2):48-62 **KING JC** *A Comparative Analysis of the Chromosomes of the Guarani Group of Drosophila*	14	3
0	5 1947 EVOLUTION 1(1-2):63-68 **AMADON D** *Ecology and the Evolution of Some Hawaiian Birds*	16	4
2	6 1947 EVOLUTION 1(1-2):69-78 **EPLING C** *Natural Hybridization of Salvia-Apiana And S-Mellifera*	26	0
0	7 1947 EVOLUTION 1(1-2):79-88 **STEBBINS GL; MATZKE EB; EPLING C** *Hybridization in a Population of Quercus-Marilandica and Quercus-Ilicifolia*	21	6
0	8 1947 EVOLUTION 1(1-2):89-102 **HEUTS MJ** *Experimental Studies on Adaptive Evolution in Gasterosteus- Aculeatus L*	92	9
0	9 1947 EVOLUTION 1(1-2):103-110 **SPENCER WP** *Genetic Drift in a Population of Drosophila-Immigrans*	5	2
0	10 1947 EVOLUTION 1(1-2):111-112 **HERSKOWITZ IH** *A New Method of Treating Drosophila-Gametes with Chemicals*	10	0

FIG. 13. Chronologic view of all papers published in *Evolution,* 1947 TO 1998 using an earlier version of *HistCite* software.

```
FN ISI Export Format
VR 1.0
PT BOOK
AU COCHRANE, AL
TI EFFECTIVENESS AND EFFICIENCY
SO EFFECTIVENESS AND EFFICIENCY
LA ENGLISH
DT BOOK
NR 80
SN 1 85315 394 X
PU THE ROYAL SOCIETY OF MEDICINE PRESS LIMITED
CR      MCKEOWN T, 1966, BLACKWELL SCIENTIFIC PUBLICATIONS, P3
        DANIELS M, 1952, BR-MED-J, V1, P1162
        PALMER JW, 1965, BR-J-PREV-SOC-MED, V19, P18
        COCHRANE AL, OXFORD UNIVERSITY PRESS
        GALLOWAY TMCL, 1963, MED-OFFR, V109, P232
        SAUNDERS J, 1970, BR-J-PREV-SOC-MED, V24, P187
        ROYAL COLLEGE OF PHYSICIANS, 1971, SMOKING AND HEALTH NOW. A REPORT
        OF THE ROYAL  COLLEGE OF PHYSICIANS LONDON- PITMAN MEDICAL AND
        SCIENTIFIC
        MATHER HG, 1971, BR-MED-J, V3, P334
        KNATTERUD GL, 1971, J-AM-MED-ASS, V217, P777
        ELWOOD PC, 1967, BR-MED-J, V4, P714
        WATERS WE, 1970, BR-MED-J, V2, P325
        BREAST CANCER SYMPOSIUM, 1969, BR-J-SURGERY, V56, P782
        COCHRANE AL, 1971, BR-J-PREV-SOC-MED, V25, P147
        FARQUHARSON EL, 1955, LANCET, V2, P517
        WEDDELL J, 200X, TO BE PUBLISHED
        LOGAN RFL, 1971, BR-MED-J, V2, P519
        HEASMAN MA, 1971, BR-MED-J, V1, P495
        BLODGET J, 1947, SURG-GYNEC-OBSTET, V84, P716
        MORRIS D, 1968, LANCET, V1, P681
        ODDIE JA, 1971, LANCET, V2, P308
        MEDAWAR PB, 1967, THE ART OF SOLUBLE (LONDON-MEUTHEUEN)
        COCHRANE AL, 1971, BR-MED-BULL, V27, P3
        MEDICAL RESEARCH COUNCIL. (1966). QUESTIONNAIRE ON RESPIRATORY
        SYMPTOMS AND INSTRUCTIONS FOR ITS USE (DAWLISH, DEVON: WJ HOLMAN)
        ROSE GA, 1962, BULL-WLD-HLTH-ORG, V27, P645
        FLETCHER CM, 1952, PROC-R-SOC-MED, V45, P577
        LEWIS SM, 1969, BR-MED-J, V4, P253
        ELWOOD PC, 1966, BR-J-RADIOL, V39, P587
        MIALL WE, 1958, CLIN-SCI, V17, P409
TC 802
BP 1
EP 103
PG 103
JI EFFECTIVENESS EFFICI
PY 1972
VL 1
IS 1
PI UNITED KINGDOM
J9 EFFECTIVENESS EFFICI
PA 1 WIMPOLE STREET, LONDON, WIM 8AE
ER
```

FIG. 14. Pseudosource record for Cochrane book, *Effectiveness and Efficiency*.

Let's turn now to an entirely different application of the *Historiographer* process to the topic of "gene flow." This is a subject of considerable interest to one of us (Pudovkin). Instead of a cited reference search, we conducted a general search in *WoS* using the simple term "gene flow" and limited the search to title words. Six hundred papers were published on this topic between 1974 and 2001 (Fig. 8). Of these, 29 papers were locally cited 10 or more times (Fig. 9), and were used to create a historiograph of the field. Thus, we are presented with a graphical description of the structure of this topic in Figure 10. Note that each rectangular node is hotlinked to the full source entry (Fig. 11). Node #3 is Richard May's 1975 paper in the *American Naturalist*.

Gene flow is an important problem in population genetics. Gene flow maintains the unity of a species: the less the gene flow, the stronger the genetic differentiation among separate geographic populations. The intensity of gene flow is very difficult to measure directly. Many indirect estimation procedures have been suggested. In 1977 Mongomery Slatkin (now at UC Berkeley) published the first paper in a series of papers on the subject (16. Slatkin, 1977) (Fig. 12). Later he suggested using measures of allelic differences among populations to estimate the gene flow. These highly cited papers are cited both globally and locally (see nodes 37, 71, 76, 121, 153). The 1980s were a time of great allozymic research activity, and allele frequency data were very easy to get. Later, in 1992, he and coauthors Hudson and Maddison (node 226) suggested the use of DNA-sequence data to estimate gene flow. This paper has not yet been cited very often because practical implementation of the approach is not easy.

The gene flow example is a work in progress because we must now correct errors or variations in cited references and feed the corrected versions into the main collection. We must also add outer references to the main collection that are perceived as important for inclusion in the historiograph. And the collection can be further increased by including all papers that cite the early work of Slatkin or others.

In constructing the complete database for any topic, one must contemplate performing an iterative process. One must decide whether or not any of the retrieved papers in a search should be used to further expand the file. This is sometimes called chain indexing. A cited reference search can be performed on any of the early citing papers to find additional papers that should be added to the original marked list. There is a wide variation in the degree of relatedness of citing papers. Some may cite many papers in the network. Others may cite only the one used for the cited reference search. Those that are heavily coupled bibliographically to the existing file should be included in the ultimate historio-

Outer nodes Missing links? Journal list All-Author list

1971-2001 Citations to A.L.Cochrane's 'Effectiveness and Efficiency'
Nodes: 835
Sorted by **year, journal, volume, page**.

Cited nodes	Nodes / Authors	GCS	LCS
0	0 1972 EFFECTIVENESS AND EFFICIENCY 1(1):1-103 **COCHRANE AL** *Effectiveness And Efficiency*	802	834
1	1 1972 INTERNATIONAL JOURNAL OF EPIDEMIOLOGY 1(4):315-318 **HETZEL BS** *Implications of Health Indicators*	6	0
1	2 1972 INTERNATIONAL JOURNAL OF EPIDEMIOLOGY 1(4):361-368 **ROSSER RM; WATTS VC** *Measurement of Hospital Output*	88	0
1	3 1972 INTERNATIONAL JOURNAL OF HEALTH SERVICES 2(4):525-529 **WHITE KL** *Teaching Epidemiologic Concepts As Scientific Basis for Understanding Problems of Organizing and Evaluating Health Services*	1	0
1	4 1972 JOURNAL OF BIOSOCIAL SCIENCE 4(4):490-494 **RICHARDS MP** *Introduction to Study of Man - Young,JZ*		0
1	5 1972 MILBANK MEMORIAL FUND QUARTERLY 50(4):17-40 **WHITE KL** *Health Care Arrangements in United-States – AD 1972*	6	1
1	6 1972 NEW ENGLAND JOURNAL OF MEDICINE 287(2):100-& INGELFIN.FJ *Randomized Clinical Trial*	34	0
1	7 1972 NEW ENGLAND JOURNAL OF MEDICINE 287(4):186-& **LISTER J** *By London Post-Effectiveness and Efficiency in Health Care*		0
1	8 1972 NEW ENGLAND JOURNAL OF MEDICINE 287(22):1125-& **SANAZARO PJ; MAGLOTT DB; ROBERTS JS; GOLDSTEI.RL; MCALLIST.JW** *Research and Development in Quality Assurance - Experimental Medical Care Review Organization Program*	39	3
1	9 1972 NEW ENGLAND JOURNAL OF MEDICINE 287(24):1223-& **WHITE KL; GAUS CR; MURNAGHA.JH** *Technology And Health Care*	11	2
1	10 1972 SCANDINAVIAN JOURNAL OF CLINICAL & LABORATORY INVESTIGATION 29():36-& **WOOD PHN** *Connective Tissue Diseases and Community*		

FIG. 15. Chronologic historiographic display of citations to Cochrane's book, *Effectiveness and Efficiency*.

0

Author(s)	COCHRANE AL
Title	EFFECTIVENESS AND EFFICIENCY
Journal	EFFECTIVENESS AND EFFICIENCY 1(1):1-103
Year	1972
Type	BOOK
Address	
Abstract	
WoS CS	802
LCS	834
cites	0
CR[80]	MCKEOWN T, 1966, BLACKWELL SCIENTIFIC PUBLICATIONS, P3 DANIELS M, 1952, BR-MED-J, V1, P1162 OFFICE OF HEALTH ECONOMICS, 1971, OHE PUBLICATION, V36 PALMER JW, 1965, BR-J-PREV-SOC-MED, V19, P18 COCHRANE AL, OXFORD UNIVERSITY PRESS GALLOWAY TMCL, 1963, MED-OFFR, V109, P232 SAUNDERS J, 1970, BR-J-PREV-SOC-MED, V24, P187 ROYAL COLLEGE OF PHYSICIANS, 1971, SMOKING AND HEALTH NOW. A REPORT OF THE ROYAL COLLEGE OF PHYSICIANS LONDON- PITMAN MEDICAL AND SCIENTIFIC MATHER HG, 1971, BR-MED-J, V3, P334 UNIVERSITIES GROUP DIABETES PROGRAM (1970) A STUDY OF THE EFFECTS OF HYPOGLYCEMIC AGENTS ON VASCULAR COMPLICATIONS IN PATIENTS WITH ADULT-ONSET DIABETES II MORTALITY RESULTS DIABETES KNATTERUD GL, 1971, J-AM-MED-ASS, V217, P777 ELWOOD PC, 1967, BR-MED-J, V4, P714 WATERS WE, 1970, BR-MED-J, V2, P325 BREAST CANCER SYMPOSIUM, 1969, BR-J-SURGERY, V56, P782 COCHRANE AL, 1971, BR-J-PREV-SOC-MED, V25, P147 FARQUHARSON EL, 1955, LANCET, V2, P517 WEDDELL J, 200X, TO BE PUBLISHED LOGAN RFL, 1971, BR-MED-J, V2, P519 HEASMAN MA, 1971, BR-MED-J, V1, P495 BLODGET J, 1947, SURG-GYNEC-OBSTET, V84, P716 MORRIS D, 1968, LANCET, V1, P681 ODDIE JA, 1971, LANCET, V2, P308 MEDAWAR PB, 1967, THE ART OF SOLUBLE (LONDON-MEUTHEUEN) COCHRANE AL, 1971, BR-MED-BULL, V27, P3 MEDICAL RESEARCH COUNCIL. (1966). QUESTIONNAIRE ON RESPIRATORY SYMPTOMS AND INSTRUCTIONS FOR ITS USE (DAWLISH, DEVON: WJ HOLMAN) ROSE GA, 1962, BULL-WLD-HLTH-ORG, V27, P645 FLETCHER CM, 1952, PROC-R-SOC-MED, V45, P577 LEWIS SM, 1969, BR-MED-J, V4, P253 ELWOOD PC, 1966, BR-J-RADIOL, V39, P587 MIALL WE, 1958, CLIN-SCI, V17, P409 COCHRANE AL, 1968, ANN-HUM-GENET, V32, P195

FIG. 16. Full source entry for node #0, Cochrane's book *Effectiveness and Efficiency*

graph. Those that are only tangentially connected need not be included. A paper with 100 cited references may have cited only one paper in the nodal network. All the other 99 will go into the list of outer references. If none of the other papers in the collection have cited any of these papers, then they will all turn up at the bottom of the outer reference list when sorted by citation frequency. You can establish a cutoff threshold for selecting only those items cited above the threshold. Thus, the only limitation on expanding search results by chain citation indexing, or by expanding the vocabulary, is the number of items that must be processed by the program. Collections up to 5,000 source items have been easily accommodated, such as the entire file of papers published in the journal *Evolution from 1947 to 1998* (http://garfield.library.upenn.edu/histomp/evol45-98/) (Fig.13).

In the early development of a topic, using unlimited or unselective chain indexing is not dangerous. But when one encounters citing papers, which are in turn highly cited, then a decision must be made as to whether or not to proceed with chaining. Thus, if the classic Lowry method for protein determination (Lowry, Rosebrough, Farr, & Randall, 1951) or some other method paper were encountered, we would have to decide whether to eliminate it from the collection. One way to temper a chain search is to look for cocitation between the highly cited paper and at least one or more of the papers in the core bibliography. What is the probability that a nonrelevant highly cited paper will have cited two or more papers on your topic? For example, it is not likely that the recent hot papers by Lander and Venter (Lander et al., 2001; Venter et al., 2001) on sequencing the human genome will have cited two or more papers on bibliographic coupling. It is highly likely that papers like the Lowry method will turn up in the outer references. But that will not cause a cascade of additional citing papers unless it is brought into the main bibliography, as would be the case in a history of protein development.

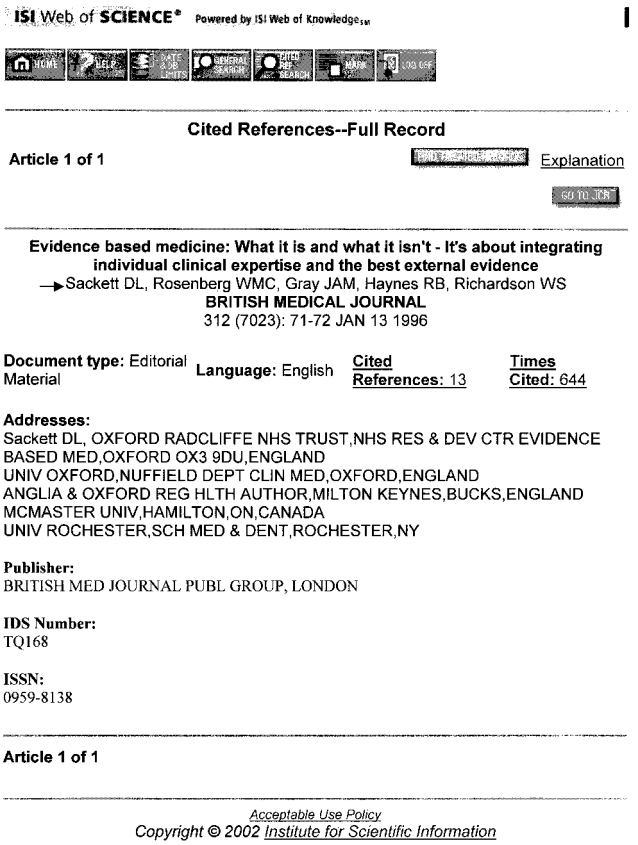
We have perhaps unduly stressed the potential errors and variations in cited references. From our experience to date the very first output of a topical search will identify the most significant papers on that topic. *HistCite* enables you to provide a more precise accounting of the citation frequencies involved. However, as was shown in the example of Marshakova, some significant works may be turned up by standardizing the variations in journal abbreviations and pagination.

Inevitably, one needs to add source documents to the main bibliography that have not been processed as sources in the *WoS*. To do this, one must manually create a source document. Consider the source record we created for the 1972 book "Effectiveness and Efficiency" by Archie Cochrane (Fig. 14). This book was cited in over 800 papers.

Web of Science location:
Cited references outside of this network.
Total: 27641 (top 100 shown).

LCS	Reference
32	SACKETT DL, 1996, BRIT MED J, V312, P71 WoS
30	ANTMAN EM, 1992, JAMA-J AM MED ASSOC, V268, P240 WoS
24	POWLES J, 1973, SCI MED MAN, V1, P1 WoS
24	CHALMERS I, 1989, EFFECTIVE CARE PREGN WoS
21	BUNKER JP, 1977, COSTS RISKS BENEFITS WoS
20	DONABEDIAN A, 1966, MILBANK MEM FUND Q, V44, P166 WoS
20	LALONDE M, 1974, NEW PERSPECTIVE HLTH WoS
20	MATHER HG, 1971, BRIT MED J, V3, P334 WoS
19	SCHWARTZ D, 1967, J CHRON DIS, V20, P637 WoS
17	MULROW CD, 1987, ANN INTERN MED, V106, P485 WoS
17	HART JT, 1971, LANCET, V1, P405 WoS
15	COCHRANE AL, 1979, MED YEAR 2000, P1 WoS
15	DICKERSIN K, 1994, BRIT MED J, V309, P1286 WoS
15	MCKEOWN T, 1976, MODERN RISE POPULATI WoS
15	MCKEOWN T, 1976, ROLE MEDICINE DREAM WoS
14	SACKETT DL, 1991, CLIN EPIDEMIOLOGY BA WoS
14	SACKETT DL, 1997, EVIDENCE BASED MED P WoS
13	NAYLOR CD, 1995, LANCET, V345, P840 WoS
13	MATHER HG, 1976, BRIT MED J, V1, P925 WoS
13	MCKEOWN T, 1979, ROLE MED WoS
13	DAVIDOFF F, 1995, BRIT MED J, V310, P1085 WoS
13	*EV BAS MED WORK G, 1992, JAMA-J AM MED ASSOC, V268, P2420 WoS
12	*UK DEP HLTH SOC S, 1976, PRIOR HLTH PERS SOC WoS
12	BEGG C, 1996, JAMA-J AM MED ASSOC, V276, P637 WoS
12	LAU J, 1992, NEW ENGL J MED, V327, P248 WoS
12	ARROW KJ, 1963, AM ECON REV, V53, P941 WoS
11	MCKEOWN T, 1976, ROLE MED WoS
11	PETO R, 1976, BRIT J CANCER, V34, P585 WoS
11	SCHULZ KF, 1995, JAMA-J AM MED ASSOC, V273, P408 WoS
11	MCKEOWN T, 1976, ROLE MED DREAM MIRAG WoS
11	RUTSTEIN DD, 1976, NEW ENGL J MED, V294, P582 WoS
11	ILLICH I, 1975, MEDICAL NEMESIS EXPR WoS
11	SACKETT DL, 1985, CLIN EPIDEMIOLOGY BA WoS

FIG. 17. Outer references for historiographic file on Cochrane's book.



ISI Web of SCIENCE® Powered by ISI Web of Knowledge℠

Cited References--Full Record

Article 1 of 1 Explanation

Evidence based medicine: What it is and what it isn't - It's about integrating individual clinical expertise and the best external evidence
→Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS
BRITISH MEDICAL JOURNAL
312 (7023): 71-72 JAN 13 1996

Document type: Editorial Material Language: English Cited References: 13 Times Cited: 644

Addresses:
Sackett DL, OXFORD RADCLIFFE NHS TRUST,NHS RES & DEV CTR EVIDENCE BASED MED,OXFORD OX3 9DU,ENGLAND
UNIV OXFORD,NUFFIELD DEPT CLIN MED,OXFORD,ENGLAND
ANGLIA & OXFORD REG HLTH AUTHOR,MILTON KEYNES,BUCKS,ENGLAND
MCMASTER UNIV,HAMILTON,ON,CANADA
UNIV ROCHESTER,SCH MED & DENT,ROCHESTER,NY

Publisher:
BRITISH MED JOURNAL PUBL GROUP, LONDON

IDS Number:
TQ168

ISSN:
0959-8138

Article 1 of 1

Acceptable Use Policy
Copyright © 2002 Institute for Scientific Information

FIG. 18. *WoS* source record for D.L. Sackett's supercited 1996 paper on evidence-based medicine.

Following the *WoS* source style, we created a pseudojournal entry for the book itself. Eighty references are cited in the book. Each of the 80 cited references in the book is tagged CR. Each comma-delimited CR line contains author, year, journal or book, volume and page. Every volume and page must be preceded by V and P.

The first part of the *HistCite* collection of papers that cite Cochrane's classic on evidence-based medicine is shown in Figure 15. Note that the GCS is lower than the LCS. This is because we edited 32 citing records that had cited a variation in pagination or whatever. GCS is the original *WoS* score. The source record for Node #0 is shown in Figure 16, and the list of outer references is shown in Figure 17. Anyone familiar with this field will not be surprised that the classic work of Dave Sackett turns up. Its WoS entry is shown in Figure 18.

When the *HistCite* collection is sorted by Local Citation Score, the 1992 work of Iain Chalmers comes to the top (Fig. 19). The broad scope of this field is demonstrated when we sort by GCS (Fig. 20) and see the highly cited work of Wennberg and others.

At the suggestion of Chaomei Chen, it was decided to test the utility of *HistCite* on the topic of "information visualization." We approached his request from the perspective of a reference librarian who is asked to find the key works on this topic. Creating a bibliographic file based on key words rather than cited authors or papers can be problematic considering the inherent ambiguity in natural language. The specific term "information visualization" is actually of recent vintage. A search of the entire 56 year *Web of Science* using "inform* visuali*" produces only 147 hits. When the search is limited to the title word field, only 50 papers were found. The oldest paper was published in 1993 (Fig. 21). However, the phrase "visualization of information" was used as early as 1977 by two Russian authors (Fig. 22). Indeed, expanding the search query to "inform* same visuali*," produces another 1000 papers published from 1985 to the present. Nevertheless, we have used the smaller sample of 147 papers to create a sample *HistCite* file (Fig. 23) based on the two variant spellings for the phrase *information visualization* or *information visualisation*. The first three entries in Figure 23 were retrieved in a search of the pre-1980 literature, which was not limited to title words. Papers in gray indicate that they neither cite nor are they cited by any other papers in the collection.

Performing a keyword search is counter to our preference for searching by cited articles or authors. Nevertheless, the natural language search has indeed been successful. By sorting on LCS we find node #4 is the most-cited paper in the bibliography. It was written by Robertson, Card, and Mackinlay at Xerox (Fig. 24). So

Outer nodes Missing links? Journal list All-Author list

1971-2001 Citations to A.L.Cochrane's 'Effectiveness and Efficiency'
Nodes: 835
Sorted by **LCS**.

Cited nodes	Nodes / Authors	GCS	LCS
0	0 1972 EFFECTIVENESS AND EFFICIENCY 1(1):1-103 **COCHRANE AL** *Effectiveness and Efficiency*	802	834
1 →	499 1992 BRITISH MEDICAL JOURNAL 305(6857):786-788 **CHALMERS I; DICKERSIN K; CHALMERS TC** *Getting to Grips with Cochrane,Archie Agenda*	174	31
2	89 1975 NEW ENGLAND JOURNAL OF MEDICINE 293(5):235-241 **HIATT HH** *Protecting Medical Commons - Who Is Responsible*	181	19
1	34 1973 SCIENCE 182(4117):1102-1107 **WENNBERG J; GITTELSO.A** *Small Area Variations in Health-Care Delivery*	531	14
1	83 1975 INTERNATIONAL JOURNAL OF HEALTH SERVICES 5(4):559-571 **RENAUD M** *Structural Constraints to State Intervention in Health*	41	11
1	55 1974 LANCET 1(7868):1179-1190 **TOWNSEND P** *Inequality and Health Service*	39	8
1	481 1991 BRITISH MEDICAL JOURNAL 303(6812):1253-1254 **LIGHT DW** *Effectiveness and Efficiency Under Competition - The Cochrane Test*	15	8
1	30 1973 NEW ENGLAND JOURNAL OF MEDICINE 289(23):1224-1229 **VAYDA E** *Comparison of Surgical Rates in Canada and in England and Wales*	160	7
1	50 1974 INTERNATIONAL JOURNAL OF HEALTH SERVICES 4(3):453-470 **DRAPER P; SMART T** *Social Science and Health Policy in United-Kingdom - Some Contributions of Social Sciences to Bureaucratization of National Health Service*	15	7
1	91 1975 NORTHWESTERN UNIVERSITY LAW REVIEW 70(1):6-68 **HAVIGHURST CC; BLUMSTEIN JF** *Coping with Quality Cost Trade-Offs in Medical-Care - Role of PSROS*	42	7

FIG. 19. Cochrane file sorted by local citation source (LCS). Chalmers 1992 paper follows Cochrane book.

Outer nodes Missing links? Journal list All-Author list

1971-2001 Citations to A.L.Cochrane's 'Effectiveness and Efficiency'
Nodes: 835
Sorted by **GCS**.

Cited nodes	Nodes / Authors	GCS	LCS
0	0 1972 EFFECTIVENESS AND EFFICIENCY 1(1):1-103 **COCHRANE AL** *Effectiveness and Efficiency*	802	834
1	34 1973 SCIENCE 182(4117):1102-1107 **WENNBERG J; GITTELSO.A** *Small Area Variations in Health Care Delivery*	531	14
2	348 1983 NEW ENGLAND JOURNAL OF MEDICINE 308(24):1457-1463 **AVORN J; SOUMERAI SB** *Improving Drug Therapy Decisions Through Educational Outreach-A Randomized Controlled*	356	1
3	326 1982 SOCIAL SCIENCE & MEDICINE 16(7):811-824 **WENNBERG JE; BARNES BA; ZUBKOFF M** *Professional Uncertainty and the Problem of Supplier-Induced Demand*	246	5
2	89 1975 NEW ENGLAND JOURNAL OF MEDICINE 293(5):235-241 **HIATT HH** *Protecting Medical Commons - Who Is Responsible*	181	19
1	499 1992 BRITISH MEDICAL JOURNAL 305(6857):786-788 **CHALMERS I; DICKERSIN K; CHALMERS TC** *Getting to Grips with Cochrane,Archie Agenda*	174	31
1	30 1973 NEW ENGLAND JOURNAL OF MEDICINE 289(23):1224-1229 **VAYDA E** *Comparison of Surgical Rates in Canada, England, Wales*	160	7
1	139 1977 INTERNATIONAL JOURNAL OF HEALTH SERVICES 7(4):663-680 **CRAWFORD R** *You Are Dangerous To Your Health - Ideology and Politics of Victim Blaming*	156	3
1	368 1984 STROKE 15(6):1068-1076 **WARLOW C** *Carotid Endarterectomy - Does It Work*	148	1
3	442 1988 JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION 259(20):3027-3030 **WENNBERG JE; MULLEY AG; HANLEY D; TIMOTHY RP; FOWLER FJ; ROOS NP; BARRY MJ; MCPHERSON K; GREENBERG ER; SOULE D; BUBOLZ T; FISHER E; MALENKA D** *An Assessment of Prostatectomy for Benign Urinary-Tract Obstruction - Geographic Variations and the Evaluation of Medical Care Outcomes*	145	5

FIG. 20. Sort by global citation score (GCS) places Wennberg 1973 paper second.

ISI Web of SCIENCE℠ Powered by ISI Web of Knowledge℠

HOME | HELP | CHANGE SEARCH LIMITS | GENERAL SEARCH | CITED REF SEARCH | LOG OFF

General Search Results--Summary

Title=inform* visuali*; DocType=All document types; Language=All languages; Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=All Years; (sorted by latest date)

SUBMIT MARKS | MARK PAGE | MARK ALL **Page 5 (Articles 41 -- 50):**

[1] [2] [3] [4] [5]

Use the checkboxes to add individual articles to the Marked List. Be sure to click SUBMIT MARKS button before leaving page.

☐ BAKER MP, BUSHELL C
AFTER THE STORM - CONSIDERATIONS FOR **INFORMATION VISUALIZATION**
IEEE COMPUT GRAPH 15 (3): 12-15 MAY 1995

☐ ROBERTSON GG, CARD SK, MACKINLAY JD
INFORMATION VISUALIZATION USING 3D INTERACTIVE ANIMATION
→COMMUN ACM 36 (4): 57-71 APR 1993

SUBMIT MARKS | MARK PAGE | MARK ALL **Page 5 (Articles 41 -- 50):**

[1] [2] [3] [4] [5]

50 of 31000834 documents matched the query.

FIG. 21. *WoS* search on information visualization—oldest paper.

ISI Web of SCIENCE℠ Powered by ISI Web of Knowledge℠

HOME | HELP | CHANGE SEARCH LIMITS | GENERAL SEARCH | CITED REF SEARCH | LOG OFF

General Search Results--Summary

Topic=information and visuali*; DocType=All document types; Language=All languages; Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977; (sorted by latest date)

SUBMIT MARKS | MARK PAGE | MARK ALL **Page 1 (Articles 1 -- 3):**

[1]

Use the checkboxes to add individual articles to the Marked List. Be sure to click SUBMIT MARKS button before leaving page.

☐ MERZLIAKOV NS, IAROSLAVSKII LP
→ **VISUALIZATION** OF **INFORMATION** USING SYNTHESIZED HOLOGRAMS
DOKL AKAD NAUK SSSR+ 237 (2): 318-321 1977

☐ EDWARDSON M, GROOMS D, PRINGLE P
VISUALIZATION AND TV NEWS **INFORMATION** GAIN
J BROADCASTING 20 (3): 373-380 1976

☐ MACKAY RS
X-RAY **VISUALIZATION** AND ANALYSIS USING SPECTRAL **INFORMATION**
IRE T MED ELECTRON 7 (2): 77-79 1960

SUBMIT MARKS | MARK PAGE | MARK ALL **Page 1 (Articles 1 -- 3):**

[1]

3 of 6654992 documents matched the query.

FIG. 22. *WoS* search on information *and* visual*, 1960–1977.

Outer References Missing Links? Journal list All-Author list Citation Matrix

Inform* Visuali* papers from 1992 to April 2002
Nodes: 147
Sorted by **year, journal, volume, page**.

Cited nodes	Nodes / Authors	GCS	LCS
0	1 1992 CANADIAN JOURNAL OF INFORMATION SCIENCE-REVUE CANADIENNE DES SCIENCES DE L INFORMATION 17(2):41-48 **GRANDE S; ROBINSON D** *Multimedia and Literacy*		0
0	2 1993 ACM TRANSACTIONS ON INFORMATION SYSTEMS 11(3):266-286 **KOIKE H** *The Role of Another Spatial Dimension in Software Visualization*	8	2
0	3 1993 COGNITIVE NEUROPSYCHOLOGY 10(1):57-77 **CASEY MB; WINNER E; BENBOW C; HAYES R; DASILVA D** *Skill At Image Generation - Handedness Interacts with Strategy Preference for Individuals Majoring in Spatial Fields*	5	0
0	4 1993 COMMUNICATIONS OF THE ACM 36(4):57-71 **ROBERTSON GG; CARD SK; MACKINLAY JD** *Information Visualization Using 3d Interactive Animation*	81	17
0	5 1994 AUTOMATION AND REMOTE CONTROL 55(7):1012-1017 **KUROCHKIN IV; MALTSEV AA** *Statistical Optimization of Interaction Between the Components of the Control Loop Early in the Design of Ergatic Systems*		0
0	6 1994 FUTURE GENERATION COMPUTER SYSTEMS 10(2-3):169-172 **PARSONS DJ** *HPCN Applications in Finance*		0
0	7 1994 INFORMATION PROCESSING '94, VOL III 53():351-358 **SCHUETT D; JACOBY K; BOCIONEK S** *Information Engineering and Innovation*		0
1	8 1995 ACM TRANSACTIONS ON INFORMATION SYSTEMS 13(3):305-323 **KOIKE H** *Fractal Views - A Fracta Based Method for Controlling Information Display*		0
0	9 1995 BRITISH TELECOMMUNICATIONS ENGINEERING 14():17-25 **WALKER G** *Challenges in Information Visualization*	1	0
0	10 1995 COMPUTER GRAPHICS FORUM 14(3):C349-C360 **BENFORD S; SNOWDON D; GREENHALGH C; INGRAM R; KNOX I; BROWN C** *Vr-Vibe - A Virtual Environment for Cooperative Information Retrieval*	4	1

FIG. 23. Chronologic file of papers on inform* Visuali, 1992–2002.

Outer References Missing Links? Journal list All-Author list Citation Matrix

Inform* Visuali* papers from 1992 to April 2002

Nodes: 147
Sorted by **LCS**.

Cited nodes	Nodes / Authors	GCS	LCS
0 →	4 1993 COMMUNICATIONS OF THE ACM 36(4):57-71 **ROBERTSON GG; CARD SK; MACKINLAY JD** *Information Visualization Using 3d Interactive Animation*	81	17
3	89 2000 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS 6(1):24-43 **HERMAN I; MELANCON G; MARSHALL MS** *Graph Visualization and Navigation in Information Visualization: A Survey*	7	5
3	38 1998 COMPUTER GRAPHICS FORUM 17(2):153-165 **HERMAN I; DELEST M; MELANCON G** *Tree Visualisation and Navigation Clues for Information Visualisation*	4	4
0	68 1999 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 50(9):799-813 **SMALL H** *Visualizing Science by Citation Mapping*	22	4
0	11 1995 COMPUTER NETWORKS AND ISDN SYSTEMS 27(6):1075-1087 **MUKHERJEA S; FOLEY JD** *Visualizing the World Wide Web with the Navigational View Builder*	8	3
1	18 1996 ACM TRANSACTIONS ON GRAPHICS 15(2):121-140 **WARE C; FRANCK G** *Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions*	12	3
0	2 1993 ACM TRANSACTIONS ON INFORMATION SYSTEMS 11(3):266-286 **KOIKE H** *The Role of Another Spatial Dimension in Software Visualization*	8	2
0	47 1998 IEEE TRANSACTIONS ON COMPUTERS 47(11):1297-1309 **PAPAKOSTAS A; TOLLIS IG** *Interactive Orthogonal Graph Drawing*	3	2
0	10 1995 COMPUTER GRAPHICS FORUM 14(3):C349-C360 **BENFORD S; SNOWDON D; GREENHALGH C; INGRAM R; KNOX I; BROWN C** *VR-VIBE - A Virtual Environment for Cooperative Information Retrieval*	4	1
1	13 1995 ERGONOMICS 38(6):1184-1198 **STANNEY KM; SALVENDY G** *Information Visualization - Assisting Low Spatial Individuals with Information Access Tasks Through Visual Mediators*	9	1

FIG. 24. Sort by local citation source (LSC) for papers on inform* Visuali*.

Author(s)	ROBERTSON GG; CARD SK; MACKINLAY JD
Title	INFORMATION VISUALIZATION USING 3D INTERACTIVE ANIMATION
Journal	COMMUNICATIONS OF THE ACM 36(4):57-71
Year	1993
Type	
Address	XEROX CORP, PALO ALTO RES CTR, 3333 COYOTE HILL RD, PALO ALTO, CA 94304 USA
Abstract	
WoS CS	81
LCS	17
cites	0
CR[25]	CARD SK, 1991, P CHI 91 HUMAN FACTO, P181 CARD SK, 1983, PSYCHOL HUMAN COMPUT CONNER DB, 1992, 1992 P S INT 3D GRAP, P183 CUTTING DR, 1991, INTELLIGENT TEXT IMA, P285 CUTTING DR, 1992, P 15 ANN INT ACM SIG, P318 FAIRCHILD KM, 1988, COGNITIVE SCI ITS AP FEINER S, 1990, P UIST 90, P76 FURNAS GW, 1986, P SIGCHI86, P16 GREEN M, 1991, COMPUT GRAPH, V25, P229 HENDERSON DA, 1986, ACM T GRAPHIC, V5, P211 JOHNSON B, 1991, P IEEE VISUALIZATION, P284 MACKINLAY J, 1990, HUMAN COMPUTER INTER, V5, P145 MACKINLAY JD, 1991, P ACM SIGCHI 91 C HU, P173 MACKINLAY JD, 1990, P SIGGRAPH 90, P171 NEWELL A, 1990, UNIFIED THEORIES COG RAO R, 1992, P UIST 92, P23 ROBERTSON GG, 1991, P ACM C HUM FACT COM, P189 ROBERTSON GG, 1989, P UIST 89 2 ANN ACM, P10 SALTON G, 1983, INTRO MODERN INFORMA SHAW C, 1992, P SIGCHI, P321 SHERIDAN TB, 1984, ADV MAN MACHINE SYST, V1, P49 SPENCE R, 1982, BEHAVIOUR INFORMATIO, V1, P43 SUTHERLAND IE, 1965, MIT296 LINC LAB TECH, P329 TESLER J, 1992, FSN 3D INFORMATION L ZELTZER D, 1989, P GRAPHICS INTERFACE

FIG. 25. Full source entry for Robertson et al. gives hint of earlier work.

Outer References Missing Links? Journal list All-Author list Citation Matrix

Inform* Visuali* papers from 1992 to April 2002

Nodes: 147
Sorted by **GCS**.

Cited nodes	Nodes / Authors	GCS	LCS
0	4 1993 COMMUNICATIONS OF THE ACM 36(4):57-71 **ROBERTSON GG; CARD SK; MACKINLAY JD** *Information Visualization Using 3d Interactive Animation*	81	17
0 →	68 1999 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 50(9):799-813 **SMALL H** *Visualizing Science by Citation Mapping*	22	4
0 →	136 2001 NUCLEIC ACIDS RESEARCH 29(5):1097-1106 **SOFIA HJ; CHEN G; HETZLER BG; REYES-SPINDOLA JF; MILLER NE** *Radical SAM, a Novel Protein Superfamily Linking Unresolved Steps in Familiar Biosynthetic Pathways with Radical Mechanisms: Functional Characterization Using New Analysis and Information Visualization Methods*	16	0
1	18 1996 ACM TRANSACTIONS ON GRAPHICS 15(2):121-140 **WARE C; FRANCK G** *Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions*	12	3
1	13 1995 ERGONOMICS 38(6):1184-1198 **STANNEY KM; SALVENDY G** *Information Visualization - Assisting Low Spatial Individuals with Information Access Tasks Through Visual Mediators*	9	1
0	2 1993 ACM TRANSACTIONS ON INFORMATION SYSTEMS 11(3):266-286 **KOIKE H** *The Role of Another Spatial Dimension in Software Visualization*	8	2
0	11 1995 COMPUTER NETWORKS AND ISDN SYSTEMS 27(6):1075-1087 **MUKHERJEA S; FOLEY JD** *Visualizing the World Wide Web with the Navigational View Builder*	8	3
0	19 1996 ARTIFICIAL INTELLIGENCE REVIEW 10(5-6):477-504 **ZAVREL J** *Neural Navigation Interfaces for Information Retrieval: Are They More Than an Appealing Idea?*	8	1
0	28 1997 COMPUTERS & GEOSCIENCES 23(4):447-456 **CARTWRIGHT W** *New Media Application to the Production of Map Products*	7	0
3	89 2000 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS 6(1):24-43 **HERMAN I; MELANCON G; MARSHALL MS** *Graph Visualization and Navigation in Information Visualization: A Survey*	7	5

FIG. 26. Sort by global citation source (GCS) of inform* visual*.

far this is the earliest paper encountered that uses the term "information visualization" as a title word. Examination of the list of references cited in this paper (Fig. 25) gives a hint of earlier work.

However, when the file is sorted by Global Citation Score, the second most cited paper, Node #68, is Henry Small's 1999 paper on "Visualizing Science by Citation Mapping," published in *JASIST* (Fig. 26).

Also in Figure 26, the third paper, by Sofia, was cited by 16 papers globally. Published in *Nucleic Acids Research* in 2001, the paper concerns a novel protein superfamily and uses information visualization methods. Thus, we are thrust into the large biotechnology literature which is outside "information visualization" per se. Note that it is not linked to any of the other papers in the sample network!

Figure 27 shows the list of outer references and identifies the two earlier 1990 conference papers by GG Robertson and Mackinlay that they cited earlier.

Following these in Figure 27 is the "granddaddy" of the overall field of visualization, Edward R. Tufte's 1983 book, *The Visual Display of Quantitative Information*. This classic

has been cited hundreds of times. So has his 1990 book *Envisioning Information*. Note also that the 1983 book has been cited as E. Tufte, so this variation should be edited to show its true rank as the most-cited outer reference.

A 1992 paper by Ben Shneidermann also appears in Figure 27. Although it has only been cited eight times, most of those occur within the network. The title, "Tree Visualizations with Tree Maps," illustrates all too well the difficulty in retrieving all relevant work easily by natural language and explains why it was not found in the original search by title.

Thus, in a 15-minute exercise, including the time to do the *WoS* search, it was possible to identify the key works on this topic. When we extended this search to a larger file of over 800 papers on information visualization, it is significant that the most-cited works were essentially the same.

Of course, a citation-conscious searcher might build up the bibliography on this topic by beginning with one or more known key papers or authors such as Robertson or Tufte. By examining successive historiographic tables, we believe all roads would essentially lead to the identification of the key works in the field.

An interesting by-product of this search is worth mentioning. When the key source paper by Robertson et al. was originally processed by ISI, the first page was listed as page

ISI Web of Science location:
Cited references outside of this network.
Total: 3225 (top 100 shown).

LCS	Reference
18	ROBERTSON GG, 1991, P ACM C HUM FACT COM, P189 WoS
13	SARKAR M, 1994, COMMUN ACM, V37, P73 WoS
12	MACKINLAY JD, 1991, P ACM SIGCHI 91 C HU, P173 WoS
11	TUFTE ER, 1983, VISUAL DISPLAY QUANT WoS
10	JOHNSON B, 1991, P IEEE VISUALIZATION, P284 WoS
10	CARD SK, 1999, READINGS INFORMATION WoS
10	TUFTE ER, 1990, ENVISIONING INFORMAT WoS
9	OLSEN KA, 1993, INFORM PROCESS MANAG, V29, P69 WoS
9	SHNEIDERMAN B, 1992, ACM T GRAPHIC, V11, P92 WoS
8	CARD S, 1999, READINGS INFORMATION WoS
8	DIBATTISTA G, 1994, COMP GEOM-THEOR APPL, V4, P235 WoS
8	FURNAS GW, 1986, P SIGCHI86, P16 WoS
8	WARE C, 2000, INFORMATION VISUALIZ WoS
8	WISE JA, 1995, P IEEE S INF VIS 95, P51 WoS
7	FRUCHTERMAN TMJ, 1991, SOFTWARE PRACT EXPER, V21, P1129 WoS
7	AHLBERG C, 1994, P ACM C HUM FACT COM, P313 WoS
7	EADES P, 1984, C NUMERANTIUM, V42, P149 WoS
7	CARD SK, 1991, P CHI 91 HUMAN FACTO, P181 WoS
7	HENDLEY RJ, 1995, P 1995 INF VIS S ATL, P90 WoS
6	FAIRCHILD KM, 1988, COGNITIVE SCI ITS AP, P201 WoS
6	MISUE K, 1995, J VISUAL LANG COMPUT, V6, P183 WoS
6	TAMASSIA R, 1987, SIAM J COMPUT, V16, P421 WoS
6	REINGOLD EM, 1981, IEEE T SOFTWARE ENG, V7, P223 WoS
6	FAIRCHILD KM, 1988, COGNITIVE SCI ITS AP WoS
6	TAMASSIA R, 1989, IEEE T CIRCUITS SYST, V36, P1230 WoS
6	TUFTE E, 1983, VISUAL DISPLAY QUANT WoS
5	RAO R, 1994, P ACM SIGCHI C HUM F, P318 WoS
5	TUFTE E, 1990, ENVISIONING INFORMAT WoS
5	SHNEIDERMAN B, 1997, DESIGNING USER INTER WoS
5	LIN X, 1997, J AM SOC INFORM SCI, V48, P40 WoS
5	CHEN C, 1999, INFORMATION VISUALIS WoS
5	HEARST M, 1995, P CHI 95, P59 WoS
5	DIBATTISTA G, 1997, COMP GEOM-THEOR APPL, V7, P303 WoS

FIG. 27. Outer references for inform* visuali*.

Potentially missed citations...
30 nodes have citations that may potentially refer to other nodes.

118 1994 COMPUTER GRAPHICS FORUM 13(3):C185-C194
SCHLEICH R; DURST MJ
BEYOND WYSIWYG - DISPLAY OF HIDDEN INFORMATION IN GRAPHICS EDITORS

ROBERTSON GG, 1993, COMMUN ACM, V36, P57 may refer to 79 ROBERTSON-GG-1993-V36-P56

169 1995 COMPUTER NETWORKS AND ISDN SYSTEMS 28(1-2):109-117
KENT RE; NEUSS C
CREATING A WEB ANALYSIS AND VISUALIZATION ENVIRONMENT

ROBERTSON GG, 1993, COMM ACM, V36 may refer to 79 ROBERTSON-GG-1993-V36-P56

174 1995 ERGONOMICS 38(6):1184-1198
STANNEY KM; SALVENDY G
INFORMATION VISUALIZATION - ASSISTING LOW SPATIAL INDIVIDUALS WITH INFORMATION ACCESS TASKS THROUGH THE USE OF VISUAL MEDIATORS

ROBERTSON GG, 1993, COMMUN ACM, V36, P57 may refer to 79 ROBERTSON-GG-1993-V36-P56

182 1995 IEEE SPECTRUM 32(11):38-&
GERSHON N; EICK SG
VISUALIZATIONS NEW TACK - MAKING SENSE OF INFORMATION

ROBERTSON GG, 1993, COMMUN ACM, V36, P57 may refer to 79 ROBERTSON-GG-1993-V36-P56

252 1996 COMPUTER JOURNAL 39(10):819-836
Ingram R; Benford S
The application of legibility techniques to enhance information visualizations

BENFORD S, 1995, COMPUT GRAPH FORUM, V14, P349 may refer to 167 BENFORD-S-1995-V14-PC349

331 1997 COMPUTERS & GEOSCIENCES 23(4):489-491
Rhyne TM
Going virtual with geographic information and scientific visualization

CARTWRIGHT W, 1997, COMPUT GEOSCI, V23, PR30 may refer to 329 CARTWRIGHT-W-1997-V23-P447

352 1997 IEEE COMMUNICATIONS MAGAZINE 35(5):72-76
Catarci T
Interacting with databases in the global information infrastructure

ROBERTSON GG, 1993, COMMUN ACM, V36, P57 may refer to 79 ROBERTSON-GG-1993-V36-P56

374 1997 JOURNAL OF NETWORK AND COMPUTER APPLICATIONS 20(2):135-149
Fowler RH; Fowler WAL; Williams JL
3D visualization of WWW semantic content for browsing and query formulation

ROBERTSON GG, 1993, COMMUN ACM, V36, P57 may refer to 79 ROBERTSON-GG-1993-V36-P56

FIG. 28. Potential pagination errors in citations to Robertson 1963 paper.

56. Figure 28 shows that the "missing links" feature of our software questioned the pagination for this paper. Did Schleich (#118) in 1994 incorrectly cite page 57? Did Kent #169, Stanney #174, and Gershon #182 repeat the error? By going back into the *WoS* it was determined that there were only about 20 cited references to page 56 but about 60 to page 57. Upon examining the original journal it was seen that page 56 was a blank page. The article really began on page 57. ISI has recently corrected this error—originally made back in 1993.

In conclusion, we have demonstrated how *HistCite* has aided the task of preparing historiographic accounts of topics, especially for the last half of the 20th century. But the process has equal validity in dealing with even more contemporary files and can give the user a snapshot of the key literature of any current topic. Whether by general keyword or by cited reference search, the method identifies the key works and permits their presentation in a traditional chronologic chart or as a more easily understood visual, genealogic chart. The process aids in identifying and correcting the many errors and variations that invariably occur in the literature.

References

Garfield, E. (2001a). "From computational liguinistics to algorithmic historigraphy." Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building large-scale knowledge bases for

intelligent applications," presented at the University of Pittsburgh on September 19, 2001. Available: http://garfield.library.upenn.edu/papers/pittsburgh92001.pdf.

Garfield, E. (2001b). From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography: A citationist's tribute to Belver C. Griffith, Lazerow Lecture presented at Drexel University, Philadelphia, PA. November 27, 2001. Available: http://garfield.library.upenn.edu/papers/drexelbevergriffith92001.pdf.

Garfield, E., Pudovkin A.I., & Istomin, V.S. (2002). Algorithmic citation-linked historiography—Mapping the literature of science. Proceedings of the American Society for Information Science and Technology, 39:14–24, Annual Meeting, November 2002.

Garfield, E., Sher, I.H. , & Torpie, R.H. (1964). The use of citation data in writing the history of science. Philadelphia: The Institute for Scientific Information, Report of research for Air Force Office of Scientific Research under contract F49(638)-1256. Available: http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf.

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. American Documentation, 14, 10+.

Lander, E.S., et al. (2001). Initial sequencing and analysis of the human genome. Nature, 408, 860–921.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., & Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. Journal of Biological Chemistry, 193, 265–275.

Venter, J.C., et al. (2001). The sequence of the human genome. Science, 291, 1304–1351.