

Author Cocitation Analysis is to Intellectual Structure As Web Colink Analysis is to... ?

Alesia Zuccala

*Statistical Cybermetrics Research Group, School of Computing and Information Technology
University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1EQ, United Kingdom
E-mail: a.zuccala@wlv.ac.uk*

Author Cocitation Analysis (ACA) and Web Colink Analysis (WCA) are examined as “sister” techniques in the related fields of bibliometrics and webometrics. Comparisons are made between the two techniques based on their data retrieval, mapping and interpretation procedures, using mathematics as the subject in focus. An ACA is carried out and interpreted for a group of participants (authors) involved in an Isaac Newton Institute (2000) workshop – Singularity Theory and Its Applications to Wave Propagation Theory and Dynamical Systems – and compared/contrasted with a WCA for a list of international mathematics research institute home pages on the Web. Although the practice of ACA may be used to inform a WCA, the two techniques do not share many elements in common. The most important departure between ACA and WCA exists at the interpretive stage when ACA maps become meaningful in light of citation theory, and WCA maps require interpretation based on hyperlink theory. Much of the research concerning link theory and motivations for linking is still new; therefore further studies based on colinking are needed, mainly map-based studies, to understand what makes a Web colink structure meaningful.

1. Introduction

Cocitation analysis is a bibliometric technique that information scientists use to “map” the intellectual structure of a research field. It involves counting documents from a chosen field – paired or cocited documents, which appear frequently in the bibliographic reference lists of citing documents. Co-citation studies compile cocitation counts in matrix form and statistically scale them to capture “a snapshot at a distinct point in time of what is actually a changing and evolving structure of knowledge” (Small, 1993, p. 5).

Author Cocitation Analysis, or ACA, is specific form of cocitation analysis based on counting highly cocited pairs of oeuvres – i.e., a body of writings by the same author, or first author in collaboration (White & Griffith, 1982, p. 257). ACA was first introduced by White and Griffith (1981a; 1981b; 1982), and described in technical detail by White (1986) in terms of cocited author retrieval and by McCain (1990) in terms of cocitation mapping. Subsequent scholars, notably Persson (2001), Ahlgren, Jarneving, and Rousseau (2003), White (2003), Rousseau and Zuccala (2004), and Leydesdorff (2004) have examined the practice of ACA and have provided suggestions for addressing its methodological problems (e.g., retrieving all author versus first author data; using Pearsons r , or Salton’s Cosine as a similarity measure).

In past years, ACA studies have appeared quite frequently (e.g. Perry & Rice, 1998; Sandstrom, 1998; White, 2003; Zuccala, 2004). Scholars are invested in this technique, yet given the debate concerning how data should be retrieved and manipulated there has been little or no disagreement about how resulting maps or “intellectual structures” should be interpreted. White (1990) lists the basic elements of a co-cited author interpretation, which are generally accepted:

- Author maps reveal the “cognitive” or “intellectual structure” of a field by showing the consensus of citers as to important contributors and works.
- Author maps show who is central and who is peripheral to a field
- The maps show who is central and who is peripheral *within clusters* representing specialties or schools of thought

- The maps show broad dimensions on which clusters are arranged. Usually one of these can be interpreted as a subject dimension and “style of work” dimension. A qualitative-quantitative polarity often appears.
- The knowledgeable interpreter of a map may see much to explicate in the fine structure of author points: for example common nationality, temporal conjunctions, teacher-student relationships, collegial and co-author relationships, or common philosophical orientations (p. 103).

Web Colink Analysis (WCA), in comparison to ACA, is a relatively new technique, based on the same pairing principle as its bibliometric “sister” – the pairing of Web colinks instead of bibliographic cocitations. In ACA, cocitation retrieval constitutes the pairing of authored papers; in WCA, colink retrieval constitutes (in this study) the pairing of academic Web sites.¹ WCA may be called the “sister” technique of ACA, because it occupies a position within a subfield of bibliometrics, known as webometrics (Almind & Ingwersen, 1997; Bjornborn & Ingwersen; 2001). Bjorneborn and Ingwersen (2001) note that “webometrics displays several similarities to informetric and scientometric studies and the application of common bibliometric methods.” (p. 65). Webometric link analysis focuses on *outlinks* pointing from web pages, and *inlinks* pointing to web pages. Both “can be seen as reference and citation analyses respectively,” but “unlike the traditional paper-based citation world... the Web often demonstrates Web pages simultaneously linking to each other” (pp. 65-66).

The term *colink* in webometrics defines an instance “when two Web pages both have inlinks from a third page” (Thelwall, 2004, p.5). Data collection for a colink analysis requires the use of search engines like AltaVista or Yahoo!; however, much of the research to date has been about the collection and measurement of inlinks and outlinks, or direct link networks (e.g., Bar-Ilan, 2005, Bjorneborn, 2004; Rousseau, 1997; Smith, 2004; Thelwall & Smith, 2002; Vaughan & Thelwall, 2003). Vaughan and Thelwall’s (2003) research addresses the issue of whether or not site age and site content are inducers of links to a journal web site. Smith (2004) examines the extent to which links on the Web are analogues to citations in traditional print literature, and Bar-Ilan (2005) has produced a classification of link types in academic environments, focusing on different aspects of the source and target pages on the Web.

Colink studies, which differ from direct link studies, are now just emerging. Few have been carried out, thus the interpretation of colink maps is fairly innovative. Larson (1996) has performed a colink analysis of a set of Earth Science related Web sites and has produced an annotated map of major topical clusters relating to weather and climate, remote sensing, geography and ecology and the environment. Polanco et al. (2001) used colinks to create a clustered map of 37 European university web sites and formulated a view of the universities based on different classes or categories. Thelwall and Wilkinson’s (2004) research concerning a network of academic web domains tested whether or not indirect connections (colinks) on the Web would be stronger indicators of subject similarity than direct links. A colink map was not created in this study; but what the authors found contradicted their prediction: “high colink counts did not give a higher probability of subject similarity” (p. 66). Vaughan and You (2005) hypothesized that the number of colinks to a pair of business Web sites could be used as a measure of similarity between two companies. In a study based on 32 telecommunications companies they were able to verify their hypothesis, and show how colink data can be used to map business competitive positions.

In this paper, a detailed comparison will be made between Author Cocitation Analysis (ACA) and Web Colink Analysis (WCA) in order to assess their similarities and differences, and determine how much the latter bibliometric technique informs its webometric “sister.” Some similar features and methodological concerns are expected, but it is understood that a WCA differs from an ACA because it is based on the theory of hyperlink creation rather than citation theory. To make this technical comparison we will focus on the subject of mathematics. We will examine an ACA mapping of authors who were involved in a workshop at the Isaac Newton Institute for Research in Mathematics and compare this to a second mapping based on a colink analysis of the Web pages of selected worldwide mathematics research institutes.

2. Author Cocitation Analysis versus Web Colink Analysis

2.1. Selecting Author Names / Selecting Web Pages

According to White (1986), a successful Author Cocitation Analysis (ACA) depends on “choosing good names on which to search. They must be authors prominent enough to have been cited by other writers in journals recorded in SciSearch® or Social SciSearch®. The authors must also be related enough in other writers’ eyes to be cited together (co-cited) with some frequency” (p. 94). When selecting names for an ACA it is essential then to choose authors that are highly cited, and that may be grouped according to a suitable rationale. Typically, the set of authors in question are selected because they publish within the same subject area (e.g., White & Griffith, 1981b, authors from judgement decision research), or because they carry out research concerning an important problem (e.g., Perry & Rice, 1998, hybrid problem area of developmental dyslexia).

Further to White’s recommendation, McCain (1990) advises that “a diversified list of authors is critical for the examination of the overall structure” of an ACA map, because “it defines the scholarly landscape being mapped. If the authors are not chosen to capture the full range of variability [e.g., research topics, methodologies, or political/national orientations, etc.]...these aspects of structure cannot be demonstrated” (p. 433). Authors chosen for an ACA should be highly cited, and possess some common elements, but they should also reflect some level of variability so that the cocitation structure will lead to a meaningful interpretation.

Consider now the selection of web pages or URLs for a WCA. If high author citation counts are important to the success of an ACA, let us assume that high link counts are of similar value to a WCA. One of the benefits of relying on Web links for a colink analysis is that they are easy to create, hence generally plentiful. Within the context of the Web, it is important to consider however that many links “may be created as a technological exercise rather than for any communication function” (Thelwall (2003; 2004). High link and high colink counts are useful from a statistical perspective, but as one shifts towards the interpretative stage of a WCA, certain colinks may not be particularly meaningful. An optimistic point of view is that “the proportion of such links does not mean that link counts cannot have a meaningful interpretation, as long as the proportion of irrelevant links is not too high” (Thelwall, 2004).

Table 1 presents a list of 51 author names associated with the Isaac Newton Institute workshop: Singularity Theory and Its Applications to Wave Propagation Theory and Dynamical Systems (2000). Again, names chosen for an ACA are normally gathered as a judgement sample, emphasising high citation counts. Typically the researcher begins working with one set of authors mined from various publications, then adds or removes names with acquired knowledge of who the prominent authors are in the subject area. Sometimes this knowledge is gained through conversations with the authors themselves (e.g., Sandstrom, 1998; Zuccala, 2004), and at other times the cocitationist is familiar already with the subject literature and its prominent authors (e.g., White & McCain’s, 1998, Visualizing a discipline: an author cocitation of Information Science, 1972-1995). Given the constrained nature of this particular ACA – i.e., a single workshop event – we diverge from the norm of focusing on high citation counts and work with a pre-determined author set (note: authors will neither be added nor removed). It is expected that some of the authors from this set may not be highly cited or even highly co-cited, but that most of them will be cocited enough for a meaningful ACA.

Table 2 presents our WCA comparison list of 44 mathematics research institutes and their home page URLs. The names of all these institutes and their URLs have been extracted from a specific “institute” directory on Google (see <http://directory.google.com/Top/Science/Math/Research/Institutes/>). Approximately 80 research institutes are listed at this Google page; however only 44 were selected for the WCA. This judgement sample of 80 URLs was reduced intentionally so that the manual retrieval procedure would not be unwieldy. Each URL was chosen on the basis of McCain’s (1990) “variability” suggestion to reflect a range of differences based on geography. The institute environments operate within different countries (note: a few are from the same country) and have been tabled with idea that the WCA for this paper will serve as an exploratory study – i.e., one that is instrumental in nature rather than an official investigation.

Table 1. ACA selection of 51 authors. Workshop participants from the *Singularity Theory and Its Applications to Wave Propagation Theory and Dynamical Systems* (Isaac Newton Institute, 2000).

ANISOV, S. S.	GRIDIN, D.	ROBERTS, M.
BARYSHNIKOV, Y.	HOUSTON, K.	ROMERO-FUSTER, M. C.
BASTO GONCALVES, J.	HOVEIJN, I.	SEDYKH, V.
BOGAEVSKY, I. A.	ISHIKAWA, G.	SHAPIRO, B.
BOLIBRUCH, A.	IZUMIYA, S.	SIERSMA, D.
BROER, H. W.	JANECZKO, S.	SIMON, P.
BUONO, P.-L.	KAZARIAN, M.	TIBAR, M.
CHAPERON, M. R.	KRUGLIKOV, B.	TOKIEDA, T.
CHEKANOV, Y. V.	LAMB, J. S. W.	TROTMAN, D. J. A.
CHILLINGWORTH, D.	LERMAN, L.	URIBE-VARGAS, R.
CUSHMAN, R.	LE DUNG TRANG	VAN NOORT, M.
DAMON, J.	MAILYBAEV, A. A.	WULFF, C.
DAVYDOV, A.	MATVEEV, V. S.	ZAKALYUKIN, V. M.
DUBROVIN, B.	MONTALDI, J.	ZHILINSKII, B. I.
FERRAND, E.	NEKHOROSHEV, N.	ZHITOMIRSKII, M. Y.
GAFFNEY, T.	OSIPENKO, G.	
GAUTHIER, J.-P.	PITANGA, P.	
GORYUNOV, V.	POLYAK, M.	

2.2. Author Cocitation / Web Colink Data Retrieval

Data collection for an ACA is based on performing a Boolean “AND” search in the Dialog™ SciSearch® or Social SciSearch® citation indexes. Two authors names are specified with the intent of retrieving any paper that cites both as first or sole author: “S CA= Author A? AND CA=Author B?” Because a first or sole author analysis does not present a complete picture of an author’s contribution to an ACA structure, Rousseau and Zuccala (2004) demonstrate how a Dialog™ search based on mathematical set theory can be used to retrieve complete author cocitation data (i.e., pure author cocitations and general author cocitations). An all-author data collection procedure is extremely tedious and should ideally be automated; however, Persson (2001) has shown that all-author maps are technically feasible. If a researcher wishes to carry out the traditional form of ACA and believes that the general “aim is not to rank authors but rather to identify research themes,” then a first or sole author analysis is sufficient (Persson, 2001, p. 343). If one is interested in developing a more detailed view of how an individual has contributed to a research area over time, regardless of his/her authorship rank, then an all-author cocitation is recommended.

Given the number of author names selected for an ACA, the combination of all possible pairs can reach a maximum number of $N(N-1)/2$. For practical reasons, White (1986) advises that “it is best to keep the input set of authors relatively small” since it is obvious that “the more pairs one starts with, the more pairs one must enter (if all combinations are to be formed)” (p. 95). To assist the cocitationist with the retrieval process, a pre-composed list of author pairings can be formatted in Microsoft Word, then cut and pasted into DialogLink™, which contains a type-ahead scroll window set for automatic input to SciSearch® or Social SciSearch®.

When collecting data for a WCA, the same Boolean “AND” search strategy is required; however, unlike the ACA retrieval process, researchers cannot benefit from an automatic input module. Search engines like Google and AltaVista prohibit users from automating retrieval sessions, unless a program is used, for example, to access the Google API (Application Programmable Interface). Normally colinks cannot be extracted from Google, so AltaVista (or one of its sister search engines) is the only possible choice for a large-scale WCA, unless the researcher is prepared to spend a significant amount of time collecting link data. To extract colinks using AltaVista, the researcher must type the following string in the advanced search window of the search engine

“link:www.msri.org AND link:www.newton.cam.ac.uk.” (worldwide; in all languages). Each unique pairing of the Web pages is done separately, and should be carried out within a day or two. The data for this study were collected on October 28th and 29th, 2004.

Table 2. WCA selection of 44 International Mathematics Research Institute pages on the Web.

Pacific Institute for the Mathematical Sciences (Vancouver, Canada).	www.pims.math.ca
Mathematical Sciences Research Institute (Berkeley, CA, USA).	www.msri.org
The Isaac Newton Institute for Mathematical Sciences (Cambridge, UK).	www.newton.cam.ac.uk
School of Mathematics, Institute for Advanced Study (Princeton, NJ, USA).	www.math.ias.edu
Fields Institute For Research in Mathematical Sciences (Toronto, Canada).	www.fields.utoronto.ca
Max-Planck-Institut fuer Mathematik (Bonn, Germany).	www.mpim-bonn.mpg.de
Centre de recherches mathématiques (Montreal, Canada).	www.crm.umontreal.ca
International Centre for Mathematical Sciences (Edinburgh, UK).	www.ma.hw.ac.uk/icms
Mathematisches Forschungsinstitut Oberwolfach (Oberwolfach, Germany).	www.mfo.de
Tata Institute of Fundamental Research, School of Mathematics (Mumbai, India).	www.math.tifr.res.in
Institut des Hautes Etudes Scientifiques (Bures-sur-Yvette, France)	www.ihes.fr
Centre International des Rencontres Mathématique (Marseille, France).	www.cirm.univ-mrs.fr
Sobolev Institute of Mathematics (Novosibirsk, Russia)	www.math.nsc.ru
Stefan Banach International Mathematical Center (Warsaw, Poland)	www.impan.gov.pl/BC/index.html
Forschungsinstitut fuer Mathematik (ETH, Zurich, Switzerland)	www.fim.math.ethz.ch
Research Institute for Mathematical Sciences (Kyoto, Japan)	www.kurims.kyoto-u.ac.jp
Czech Research Consortium for Informatics and Mathematics (Czech Republic).	www.utia.cas.cz/CRCIM
Euler International Mathematical Institute (St. Petersburg, Russia)	www.pdmi.ras.ru/EIMI
Mittag-Leffler Institute Royal Swedish Academy of Sciences (Stockholm, Sweden)	www.ml.kva.se
International Centre for Pure and Applied Mathematics (Nice, France)	www-mathdoc.ujf-grenoble.fr/CIMPA
Erwin Schrödinger Institute for Mathematical Physics (Vienna, Austria)	www.esi.ac.at
Euler Institute for Discrete Mathematics and its Applications (Eindhoven, The Netherlands).	www.win.tue.nl/math/eidma
Institute for Mathematical Sciences (Singapore).	www.ims.nus.edu.sg
Centre de Recerca Matemàtica (Barcelona, Spain)	www.crm.es
Centre for Mathematical Physics and Stochastics - "MaPhySto" (Aarhus, Denmark).	www.maphysto.dk
Institute of Mathematics "Simion Stoilow" of the Romanian Academy (Bucharest, Romania).	www.imar.ro
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences (Budapest, Hungary).	www.math-inst.hu
Mathematics Research Centre (Warwick, England, UK).	www.maths.warwick.ac.uk/mrc
African Institute for Mathematical Sciences (Cape Town, South Africa).	www.aimsforafrica.org
New Zealand Institute of Mathematics and its Applications (Auckland, New Zealand).	www.nzima.auckland.ac.nz
Harish-Chandra Research Institute (India).	www.mri.ernet.in
Centro Internacional de Matemática (Portugal).	www.cim.pt
Rolf Nevanlinna Institute (Finland).	www.rni.helsinki.fi
Steklov Institute of the Russian Academy of Sciences (Moscow, Russia).	www.mi.ras.ru/index_e.html
National Academy of Sciences of Ukraine (Kyiv-4, Ukraine).	www.imath.kiev.ua/en
Centre for Mathematics and its Applications (Canberra, Australia).	www.maths.anu.edu.au/CMA
Feza Gürsey Institute (İstanbul, Turkey).	www.gursey.gov.tr
Tbilisi International Centre of Mathematics and Informatics (Georgia).	www.viam.hepi.edu.ge/others/ticmi
IPM - Institute for Studies in Theoretical Physics and Mathematics (Tehran, Iran).	www.ipm.ac.ir
School of Mathematics, Korea Institute for Advanced Study (Seoul, Korea).	www.kias.re.kr/en/programs/math.jsp
L'Istituto per le Applicazioni della Matematica e dell'Informatica (Milan, Italy).	www.iami.mi.cnr.it
Chennai Mathematical Institute (Chennai, India).	www.cmi.ac.in
Slovak Academy of Science Mathematical Institute (Bratislava, Slovak Republic).	www.saske.sk/MI
Centro de Investigación en Matemáticas (Guanajuato, Mexico).	www.cimat.mx/info_general/english.html

A point of interest when performing a WCA is that link counts on the Web are typically unstable (e.g., Bar-Ilan, 1999; Rousseau, 1999; Snyder & Rosenbaum, 1999). The Web as a data resource tends to fluctuate, thus one may expect that a colink search will yield a count of 45 one day, and a count of 56, perhaps one day later. Citation indexes, on the other hand, are more stable and reliable, but obviously different from the Web in that they do not provide “up to the minute” data. A cocitationist works with data that has taken a significant period of time to accumulate. ACA does not reflect the current publishing activities of the authors in question: it is an historical analysis of the authors’ past work. The process leading to an indexed citation, beginning with one author’s motivation for citing another author, is also a more complex and formalized procedure. Thelwall

(2003) reminds us that “a journal article may be rejected for publication due to an author’s failure to cite a relevant article” but a web page, by comparison would never be rejected due to the web author’s failure to link to a page on an external site (p. 7).

When performing a WCA, the manual retrieval of data using AltaVista is time-consuming; however, there is no problem equivalent to the first-author versus all-author retrieval problem in ACA. A Web site, by definition, is “a self-contained collection of one or more pages with a consistent theme” (Thelwall, 2004). As the colink retrieval is performed using AltaVista, the pairing process retrieves only pages with links to the home pages of sites and not the sites’ branching pages. Omitting a site’s branching pages is however not as serious as omitting authors from an ACA, where all authors should be recognized as individual research contributors. If we try to include every individual page of a web site on a WCA map, the technique is not necessarily enhanced. Homepage colink structures still support meaningful and valid interpretations because the comparison can be made at the organization level. Also, many links are targeted at the home pages of organizational Web sites (Thelwall, 2002), except in the case of research oriented links, which are more likely to be made within a Web site’s content to specific articles, researchers or projects.

The fourth issue, which is particular to ACA, but less important for a WCA is the problem of retrieving homonymous data. Homonymous web sites are not expected with the practice of WCA because web pages are registered domains: no two domains are accidentally alike. Nevertheless, Web sites can host mirror copies of others’ Web sites. For example, many universities publish copies of computing documentation and commonly use sites such as the DBLP computer science bibliography. Links to and from mirror sites are ‘redundant’ in the sense of not being directly associated with the host site owners. Dialog™ SciSearch®, by comparison, is an international multidisciplinary database for all articles, review papers, and meetings abstracts published in the fields of mathematics, chemistry, agriculture, biology, environmental sciences, clinical medicine, the life sciences, and the engineering sciences and physics. Due to the broad coverage of this database the pairing of author names – homonymous authors from the same specialty or other research specialties – could lead to irrelevant cocitation counts. Most cocitation analysts agree that it is unlikely that two or more authors with the same surname and first name initial will be active in the same research specialty, or have “twins” paired together as co-cited authors in another scientific specialty (Harter, 1986; Sandstrom, 1998; White, 1986). Harter (1986) claims that the number of such pairings is “exceedingly small” (p. 187). Likewise, White (1986) has found that “homonym discrimination breaks down only in rare cases” and provides an example search in LIS (CR=WILSON P?) which has the effect of retrieving papers citing either Patrick or Pauline Wilson in LIS (p. 95). Sandstrom (1998) also recognizes that “significant skewing can conceivably occur when two very common names are combined, but this question is one that should be addressed empirically” (p. 164, original emphasis).

And finally, the fifth point of comparison between the two research techniques requires that we give some attention to the use of author self-cocitations in ACA versus Web site self-colinks in WCA. When an author cites her own work and the work of another author in her own paper’s bibliographic reference list, she creates a self-cocitation. Likewise a Web site that links to both its own home page and that of another organization’s home page is producing what we call a site self-colink. With the practice of ACA, we typically include self-cocitation data in the retrieval and mapping procedure because it is understood that authors can and should contribute to their own cocitation profile. Not only do we want to know how authors view each other as external citers, but also how the individual authors view themselves: “my work is similar to or not similar to” author X within a research community. With the practice of WCA, the use of site self-colinks in the data retrieval and mapping procedure is not as straightforward. Some researchers might argue that it is important to exclude site self-colinks because it will ensure a “clean” data set, which will not skew the meaning of a colink map. For instance, many site self-colinks should not be counted because they occur unintentionally – Web site developers often include a return link to their site’s home page on every page of a site merely for navigation purposes. Other researchers might argue that if we do not remove site self-colinks in the retrieval process, we are not adhering to a true definition of a colink – i.e., Co-linked: when two pages both have inlinks from a third page (Bjorneborn & Ingwersen, 2001; Thelwall, 2004). Clearly, this argument depends precisely on the definition used in the research, and at present the above definition does not specify that the third linking page must be an “outside” page from another Web site. To that end, it is up to the individual researcher to provide a reason for

removing or including all site self-links or perhaps determine by comparison how both data circumstances might impact the mapping procedure. In this WCA of the international institutes site self-colinks are not removed and the decision to include them is based on two reasons: 1) we want to determine the extent to the Web site self-colinks are contributing to the international institute colink map; and 2) we consider the underlying motivations for site self-links to be an important part of the map's meaning or interpretation.

2.3. Cocitation / Colink Matrices and Summary Statistics

Figures 1 and 2 below present the cocitation and colink maps of the Singularity Theory - Wave Propagation /Dynamical Systems workshop authors and the international mathematics research institute home page URLs. Both maps have been created from separate adjacency matrices (shown in Tables 3 and 4), and submitted to SPSS-10 for a multidimensional scaling and clustering routine.

At present there is a disagreement among scholars concerning the use of Pearson's r as a similarity measure in ACA. Ahlgren, Jarneving & Rousseau, (2003) and Leydesdorff (2004) oppose the use of this coefficient, while White (2003) believes it is acceptable and "performs well enough for an ACA" (p. 1250). In light of this disagreement, we note that Pearson's r was the chosen similarity measure for this study, not just for the ACA, but also for the WCA. The author understands that Pearson's r is not statistically optimal because it fails two important tests of measurement stability, but has decided to accept White's (2003) argument that the Pearson's r coefficient produces clusters and maps very like those based on other coefficients. The resulting cocitation map should therefore not be any more or less useful for an interpretation; however, new colink research now suggests that the Pearson's r correlation coefficient may be more useful for colink mapping because it presents less distortion among point similarities and dissimilarities (Ortega-Priego & Aguillo, 2005, p. 203).

With ACA, there is also an ongoing problem concerning what the researcher should do with the diagonal cells of an ACA/WCA matrix. McCain (1990) notes that she has "experimented with an alternate approach – treating the diagonal cells as missing data," and upon examining the results she found "little difference, in mapping, clustering and factor analysis, between scaling the diagonal values...and treating them as missing data" (p. 435). Ahlgren, Jarneving & Rousseau (2003) take an alternative view, stating that "some methods for creating diagonal values are at best inelegant, and at worst completely arbitrary (p. 551). To solve the diagonal problem, the authors suggest using "the number of times an author, say AU, has been cocited with him/herself (excluding self-citations)." (p. 551). In this study, column averages were inserted into both the ACA and WCA matrix diagonals. Currently, there is no effective method for retrieving data through Dialog™, AltaVista, or Google based on Ahlgren et al's (2003) suggestion.

Table 3. ACA data matrix (partial) and summary statistics. Authors involved in the *Workshop on Singularity Theory and Its Applications to Wave Propagation Theory and Dynamical Systems* (Isaac Newton Institute, 2000).

	ANISOV SS?	BARYSHNIKOV Yuliy	BASTO GONCALVES	BOGAEVSKY IA	BOLIBRUCH A	BROER HW	BUONO P-L	CHAPERON MR	CHEKANOV Y?	CHILLINGWORTH D	CUSHMAN R	DAMON J?	DAVYDOV A?	DUBROVIN B	FERRAND E	GAFFNEY T?	GAUTHIER J-P	GORYUNOV V?
ANISOV SS?	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1
BARYSHNIKOV Yuliy?	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BASTO GONCALVES J?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BOGAEVSKY IA	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
BOLIBRUCH A	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
BROER HW	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
BUONO P-L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHAPERON MR	1	0	0	0	0	0	0	1	9	0	1	0	1	0	5	0	0	1
CHEKANOV Y?	1	0	0	0	0	0	0	9	1	0	0	1	0	0	8	0	0	1
CHILLINGWORTH D	0	0	0	0	0	0	0	0	0	1	1	6	0	0	0	2	0	0
CUSHMAN R	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
DAMON J?	0	0	0	1	0	1	0	0	1	6	0	5	2	0	0	43	0	18
DAVYDOV A?	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	3
DUBROVIN B	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	1	1
FERRAND E	1	0	0	0	0	0	0	5	8	0	0	0	0	0	1	0	0	6
GAFFNEY T?	0	0	0	0	0	0	0	0	0	2	0	43	0	0	0	3	0	2
GAUTHIER J-P	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
GORYUNOV V?	1	0	0	0	0	0	0	1	1	0	0	18	3	1	6	2	0	2

51(51-1)/2 =	1275 unique pairs
Mean cocitation rate (over 51 authors):	.62
Range of raw cocitation counts:	0 to 43 (e.g., high cocitation count between DAMON and GAFFNEY)
Range of mean cocitation counts:	0 to 4 (e.g., BASTO GONCALVES, GRIDIN, KRUGLIKOV, MAILYBAEV, TOKIEDA, VAN NOORT – low; DAMON – high).
Number of unique cocitation pairings never made:	1076 (16% connectivity ratio)
pairings made only once:	113

Table 4. WCA data matrix (partial) and summary statistics. URLs for 44 international mathematics research institutes featured on the web.

	www.pims.math.ca	www.msri.org	www.newton.cam.ac.uk	www.math.ias.edu	www.fields.utoronto.ca	www.mpim-bonn.mpg.de	www.crm.umontreal.ca	www.ma.hw.ac.uk/icms	www.mfo.de	www.math.tifr.res.in	www.ihes.fr	www.cirm.univ-mrs.fr	www.math.nsc.ru	www.impan.gov.pl/BC/index.h	www.fim.math.ethz.ch/
www.pims.math.ca	63	575	149	111	402	94	211	83	106	48	80	29	21	45	64
www.msri.org	575	91	409	307	306	257	120	174	350	68	181	82	13	62	73
www.newton.cam.ac.uk	149	409	78	186	233	210	100	243	290	57	176	80	22	90	74
www.math.ias.edu	111	307	186	54	145	171	67	95	171	62	124	32	22	38	69
www.fields.utoronto.ca	402	306	233	145	68	146	236	119	193	52	115	52	13	54	69
www.mpim-bonn.mpg.de	94	257	210	171	146	61	74	112	292	68	174	69	15	69	73
www.crm.umontreal.ca	211	120	100	67	236	74	42	65	75	36	54	18	21	31	45
www.ma.hw.ac.uk/icms	83	174	243	95	119	112	65	51	144	47	88	48	14	68	59
www.mfo.de	106	350	290	171	193	292	75	144	72	61	173	129	14	77	70
www.math.tifr.res.in	48	68	57	62	52	68	36	47	61	31	57	12	20	30	42
www.ihes.fr	80	181	176	124	115	174	54	88	173	57	52	81	23	75	73
www.cirm.univ-mrs.fr	29	82	80	32	52	69	18	48	129	12	81	22	0	38	20
www.math.nsc.ru	21	13	22	22	13	15	21	14	14	20	23	0	15	22	21
www.impan.gov.pl/BC/index	45	62	90	38	54	69	31	68	77	30	75	38	22	32	43
www.fim.math.ethz.ch	64	73	74	69	69	73	45	59	70	42	73	20	21	43	34

44(44-1)/2 =	946 possible unique pairs
Mean colink rate (over 44 Web pages):	31
Range of raw colink counts:	0 to 575 (e.g., high colink count between www.pims.math.ca and www.msri.org; low colink count between www.math.nsc.ru and www.cirm.univ-mrs.fr)
Range of mean colink counts:	7 to 93 (www.cim.pl - lowest mean colinks; www.msri.org - highest mean colinks).
Number of unique colink pairings never made:	16 (98% connectivity ratio)
pairings made only once:	9

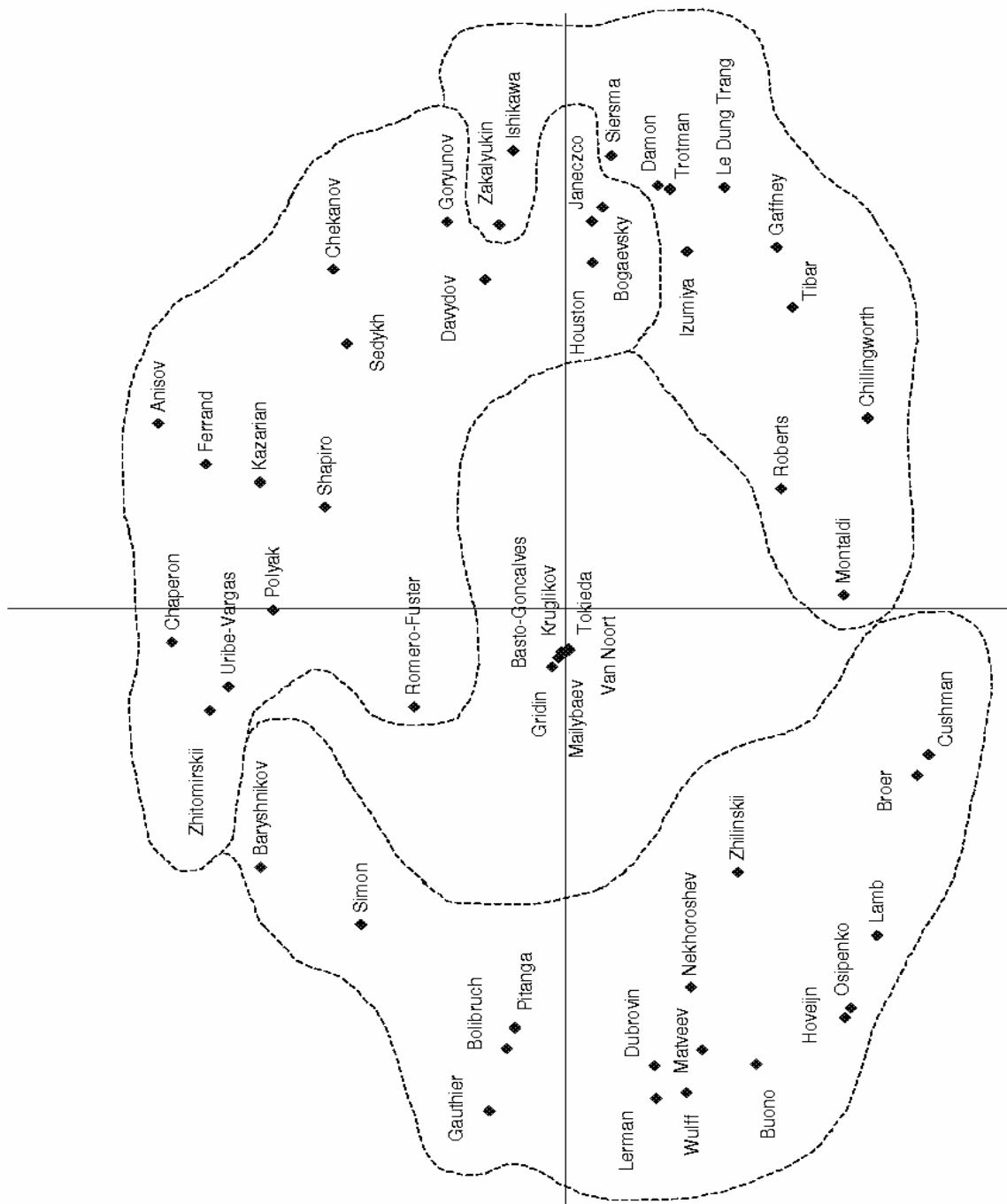


Figure 1. Author Cocitation Map of the 51 participants of the Singularity Theory and Its Applications to Wave Propagation Theory Workshop (Isaac Newton Institute, 2000).

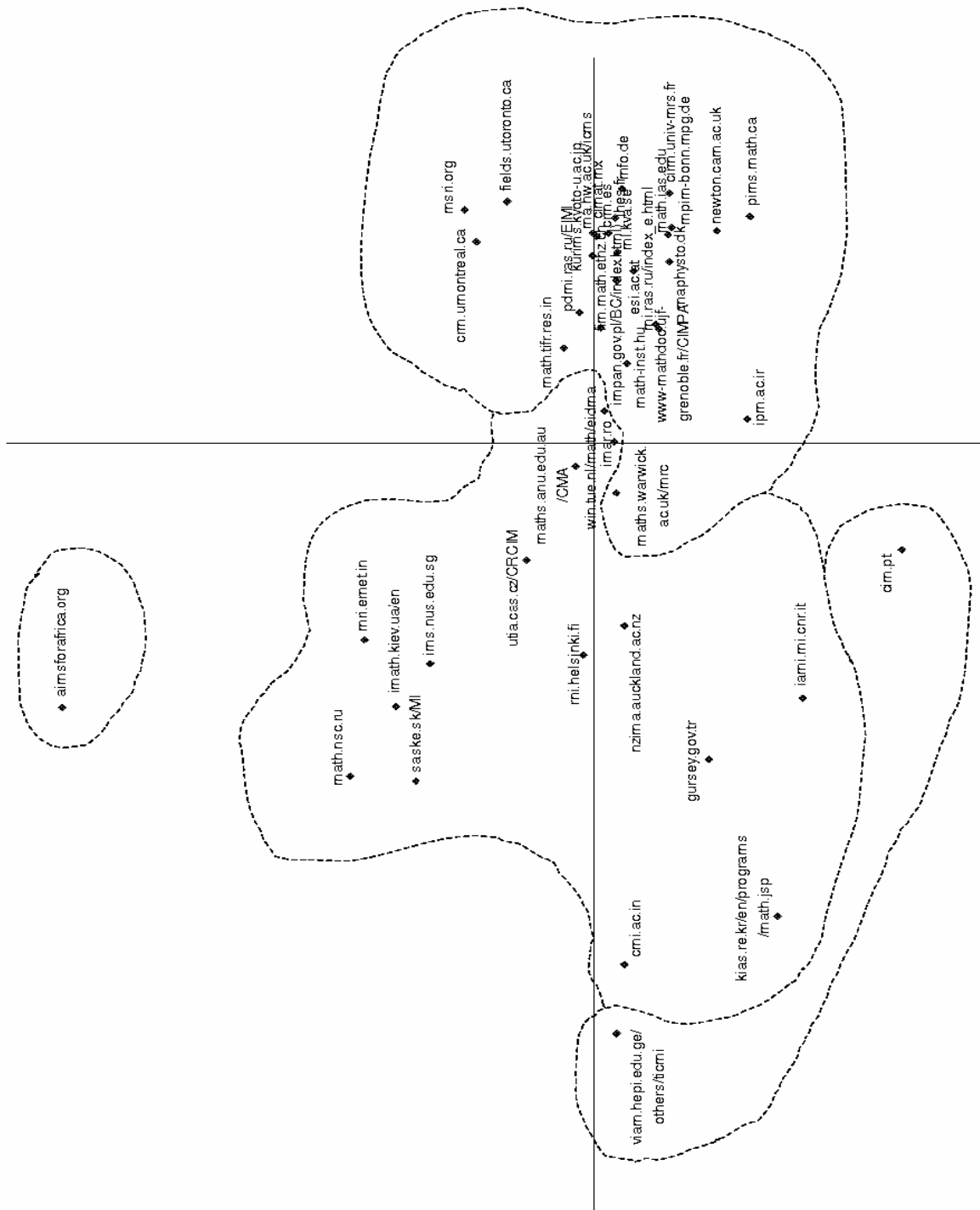


Figure 2. Web Colink Map of 44 International Mathematics Research Institutes (November, 2004).

2.4. ACA / WCA Mapping and Interpretation

2.4.1. Multidimensional scaling

A multidimensional scaling routine in ACA is based on an algorithm used by SPSS-10 to “create visual displays – maps – from proximity matrices (i.e., Pearson’s r proximities), so that the underlying structure within a set of objects can be studied” (McCain 1990, p. 437). The major output of an MDS routine in ALSCAL is a scatter plot of points or author nodes mapped in two or three-dimensional space. Generally, the closer the multidimensional scaling routine’s RSQ is to 1 (percentage of the total variance explained), the better the outcome. Also, the smaller the stress value (e.g., $< .20$) the better the data fit the model.

The ACA mapping of the Singularity Theory-Wave Propagation authors (Figure 1) resulted in a RSQ of .65849 and a stress value of .25307. Comparatively, the colink mapping of the international mathematics research institute URLs (Figure 2) resulted in a RSQ of .91082 and a stress value of .16593. When both proximity matrices were used as input to ASCAL for a three-dimensional solution, minor improvements to the RSQ and Stress values were observed (ACA 3D: RSQ=.72103; Stress=.19847 and WCA 3D: RSQ=.93863; Stress=.12385). The 3-dimensional solutions were a better fit for both the ACA and WCA data; however, we use the two-dimensional solutions for the interpretation because they are more visually accessible.

When using ACA, structural anomalies or distortions can sometimes occur when using a two-dimensional scaling solution (McCain, 1990). The same is true also for a WCA. On Figure 2, we see that the Mathematical Sciences Research Institute at Berkeley (msri.org) and the Pacific Institute of Mathematical Sciences in Canada (pims.math.ca) sit apart as distant nodal points. Here we have a colink anomaly, because the two institute Web pages, relative to all of the other pages, have the highest number of colink counts (count=575). The SPSS-10 algorithm has distorted the Web pages’ similarity; hence a three-dimensional solution to the data, using advanced mapping software, would allow us to see the two URLs positioned closer together.

Note from Figure 1 and Figure 2 that there is a shared core-periphery configuration around the individual maps’ x and y-axes. Certain authors from the ACA map (Figure 1) are located near the centre (e.g., TOKIEDA) and others occupy space around the periphery (e.g., ZHITOMIRSKII). Likewise, the institute URLs on the WCA map (Figure 2) are scattered about the axes in more or less central nodal positions: The Centre for Mathematics and Its Applications in Canberra, Australia (maths.amu.edu.au/CMA) sits in the middle, while the African Institute for Mathematical Sciences in Cape Town, South Africa (aimsforafrica.org) is decidedly peripheral.

With an ACA map, a core-periphery configuration is both familiar and expected: “points representing authors with high similarities [are] placed close together in ‘intellectual space,’ while points representing authors with high dissimilarities [are] placed farther apart (McCain, 1990, p. 438). Co-cited authors are “intellectually similar” because their citing colleagues perceive their works to be cognitively relevant (e.g., DAMON and TROTMAN on Figure 1). Sometimes the cognitive relevance is quite direct: two authors’ separate works are needed by the citing author to develop a new idea. At other times it may be less direct: authors are cocited regularly with other authors because they are well known in the subject area – i.e., citers are motivated to pay homage to pioneers (Garfield, 1965). Cronin’s (1994) view of direct versus less direct citing has been discussed previously in the literature in terms of “tiered citations.” On maps based on tiered citations, the perceived similarity between pairs of cited documents does not have to be a problem if the researcher considers “the notion that there are levels or gradations of citation” (p. 538).

Figure 1 presents a cluster of central author nodes that are similar (i.e., TOKIEDA; BASTOGONCALVES; KRUGLIKOV; VAN NOORT; GRIDIN; MAILYBAEV), but not for the usual reason of being highly cocited. Many of the authors used in this analysis, including the names listed above, possessed low citation and cocitation counts in Dialog™ SciSearch®. Due to the low counts, the multidimensional scaling routine has grouped certain names together to indicate that they are either students (with few or no publications) or newcomers to the Singularity Theory/Wave Propagation subject. These centrally placed authors are (or were, at the time of the analysis in September 2000) not as cognitively identifiable as some of the other authors on the map.

On Figure 2, the core-periphery arrangement of URLs around the x and y-axes is more difficult to assess. At a glance, we can appreciate that certain institute Web pages are colinked together more or less often, but why? *Author Cocitation Analysis is to Intellectual Structure, as Web Colink Analysis is to...*?

In “What is this link doing here” Thelwall (2003) reviews the literature on citation theory and recognizes that “other than links in e-journal articles and online copies of preprints, very few hyperlinks between sites are created as a result of a necessity on par with that for citations” (§12) Web links are categorically different from bibliographic citations; thus one of the goals of webometrics research has been to develop a link motivation theory, parallel to existing citer motivation theory (e.g., Bar-Ilan, 2005; Chu, 2003; Harries et al., 2004; Kim, 2000; Thelwall, 2001; Thelwall, 2003). At present this theory is still under development, and though it is possible to uncover a Web site creator’s impetus for creating a link, “hyperlinking behaviour frequently results from a complex interplay of motivations” (Kim, 2000, p. 887).

Thelwall’s (2003) link motivation research, which we use for this particular analysis, presents “a fine-grained process of differentiating between creation motivations for links in academic Web sites” (*abstract*). Direct links to academic sites occur for the following reasons:

- 1) *general navigational* (i.e., links to a wide variety of non-subject specific information)
- 2) *ownership link* (i.e., those that acknowledge authorship, co-authorship, ownership or co-ownership of the host Web page(s) or associated project)
- 3) *social links* (i.e., those created with the apparent primary purpose of reinforcing social ties)
- 4) *gratuitous links* (i.e., those without any discernable communication motivation behind their creation).

Based on Thelwall’s (2003) work, we might hypothesize that the colink structure of the international mathematics research institutes is social in nature (i.e. representing a social structure). Unless we visit and examine several of the colinking pages, this hypothesis cannot be verified. Let us focus now on the practice of cluster analysis in ACA, to see if a CLUSTER routine can be more helpful to our WCA interpretation.

2.4.2. Clustering

In ACA, the use of CLUSTER in SPSS-10 aids the interpretive process by creating an enhanced view of an intellectual mapping, showing how certain authors group together and others do not. Normally a hierarchical agglomerative approach to this analysis is used, although there are in fact “150 specific methods” to choose from in CLUSTER. Each may be used “to group objects, people, countries, or other entities on the basis of shared attributes” (McCain, 1990, p. 437). With an agglomerative routine, there is a “bottom-up building” process. Authors and, in this case, URLs are joined together gradually to create smaller clusters, then subsequently joined again to create larger clusters, until there is a complete linkage. The resulting SPSS display is a dendrogram, which assists in determining cluster cut-off points.

On both Figure 1 and Figure 2, results of an SPSS-10 CLUSTER routine have been added to the maps with the inclusion of dashed boundary lines drawn around the authors and the URLs (note: the cut-off points from the dendrogram have been selected at a stage close to complete-linkage). With the cluster groups identified, the researcher has an opportunity to produce a label, and reflect on an overall reason as to why they fit together: *Is the structure “puzzle-like” in its configuration, or more akin to a group of distant “blobs”?*

Labelling the clusters on the Singularity Theory-Wave Propagation theory map (Figure 1) is a straightforward process. Citation theory informs us that authors mapped in close proximity are “intellectually similar;” therefore clustered authors are also similar, but from a more precise topic or subtopic perspective.

The authors featured on the ACA map (Figure 1) are mathematicians. When mathematicians publish new research, their papers are indexed according to the American Mathematics Society’s

Subject Classification System. To create cluster labels, we compile a list of the clustered authors' names and perform a search in DialogTM MathSci® for all of their papers (Search AU=BARISHNIKIV, Y? OR AU=SIMON, P? OR AU=BOLYBRUCH, A? ... etc). Once the papers are retrieved, we RANK them according to their assigned classification codes (RANK DC). With the RANK command in DialogTM we can determine the top four/five topics or description code areas that the clustered authors share in common. Figure 3, below, illustrates this process.

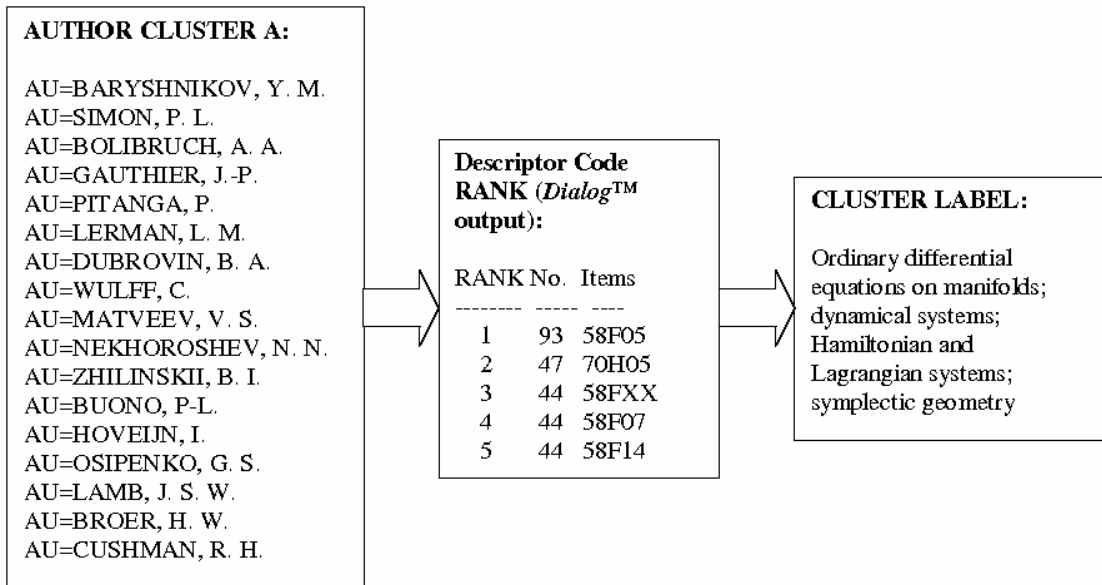


Figure 3. Descriptor code ranking, using DialogTM MathSci and clustered authors.

On the ACA Map (Figure 1), the labelled clusters fit together in an oval-like pattern. A birds-eye view of the authors shows that they are not well integrated: some of the mathematicians are aligned with Singularity Theory research and its associated topics, while others are more involved in research concerning Wave Propagation Theory and/or Dynamical Systems.

The aim of the Singularity Theory-Wave Propagation workshop was to inspire research interaction among the invited participants (authors). The institute environment provided the authors with a collaborative context for sharing information and developing new ideas. During the workshop's weeklong period (in October, 2000), the information sharing process had only just begun: the ACA map captures a distinct point in time when a select group of researchers were meeting together with different research histories. In the future, the map may be used again, as a tool for assessing the workshop's aim and outcome. The ACA may be reiterated in five to ten years time, including new cocitation counts, to find out if there is a noticeably altered configuration. Perhaps the oval pattern will draw in slightly to reveal a more "puzzle-like" structure. Perhaps it might not. In any case, the general value of the ACA map is that it can be useful for predicting and examining knowledge growth.

Look closely now at Figure 2: the WCA map of the international mathematics research institutes. For each of the clusters shown here, there is no ranking and labelling procedure similar to that performed on DialogTM MathSci®. To label or interpret our Web colink clusters, one possible strategy is to compile lists of target pages, which lead to colinks and examine the colink situations individually for common themes. In this study, we use a specially designed link extraction program, to collect the first 50 results (target pages) of an AltaVista colink retrieval session. Table 5 presents a sample list for The Fields Institute and the Mathematical Sciences Research Institute at Berkeley (i.e., link:www.fields.utoronto.ca AND link:www.msri.org).

Table 5. Link Extraction list of the first 50 results in AltaVista for link:www.fields.utoronto.ca AND link:www.msri.org.

1	http://www.math.upenn.edu/MathSources.html	Electronic Sources for Mathematics
2	http://orion.math.uwaterloo.ca/~hwolkowi/intrstsites.html	Henry Wolkowicz: Interesting Sites Related to Mathematics
3	http://camel.math.ca/Services/web.html	Other Mathematics Service Providers
4	http://www.fields.utoronto.ca/aboutus/IMSI.html	Fields Institute - Aboutus/IMSI
5	http://www.snark.org/~pjhughes/math.htm	The Telson Spur: Field Nodes -- Ground (8): Mathematics
6	http://www.math.psu.edu/MathLists/Institutes.html	PSU Math - Other Math Sites - Institutes and Centers
7	http://www.combinatorics.org/Conferences	Forthcoming Conferences in Combinatorics
8	http://www.math.wisc.edu/~kurtz	Tom Kurtz Home Page
9	http://www.ams.org/mathweb/mi-inst.html	Math on the Web: Institutions
10	http://www.ictp.trieste.it/www_users/math/mathlinks.html	ICTP Mathematics Group: Useful Links
11	http://www.math.tamu.edu/~Ken.Dykema	Ken Dykema
12	http://www.lib.virginia.edu/science/guides/s-math.htm	SEL Resources: Mathematics Research Guide
13	http://www.cohomology.com/	www.cohomology.com
14	http://www.math.tamu.edu/directory/mathlinks.html	Mathematical Links
15	http://www.maths.lth.se/links/mathlink.html	Mathematics on the Web
16	http://www.math.miami.edu/~ruan	Dr. Shigui Ruan's Homepage
17	http://www.svpal.org/~lsch	Larry's Home Page
18	http://www.math.ias.edu/mathlinks.html	Mathematics Resources
19	http://www.math.clemson.edu/societies.html	Mathematical Groups
20	http://www.desargues.univ-lyon1.fr/home/wagner/fowae.html	Frank Wagner
21	http://www.auckland.ac.nz/libr/math/mathorg.htm	People and Organizations - Mathematics - LEARN - The University of Auckland Library
22	http://www.math.uiowa.edu/mathlinks.html	Math Links - Dept of Mathematics, CLAS, University of Iowa
23	http://www.math.utk.edu/~guan	Bo Guan: Home Page
24	http://www.msccs.dal.ca/~ruan	Dr. Shigui Ruan's Stuff
25	http://www.math.uio.no/~bratteli/links.html	BRATTELI, PRIVAT
26	http://www.ub.es/logica/casanovas/i.html	Enrique Casanovas
27	http://www.lib.uchicago.edu/e/su/math	Mathematics
28	http://www.math.toronto.edu/henrique	Henrique's Home Page
29	http://math.arizona.edu/resources	Arizona Mathematics Resources
30	http://www.math.duke.edu/~wiletski/scamr.html	Sci. Comp. and Applied Math Resources
31	http://www.math.utah.edu/~kapovich/ma.html	Mathematics on the Web
32	http://www.math.wisc.edu/~ram/math.html	Mathematics
33	http://www.math.wm.edu/~shij/math.html	www.math.wm.edu/~shij/math.html
34	http://ukrgap.exponenta.ru/UkrGAP/links.htm	Bookmarks
35	http://www.abc.se/~m9847/matre/center.html	Mathematical Resources: Centers and Institutes (Math Links by Bruno Kevius)
36	http://www.fields.utoronto.ca/programs/scientific/04-05/arthurconf/	Fields Institute - Conference on automorphic forms and the trace formula
37	http://www.ipam.ucla.edu/links.aspx	IPAM - Links Page
38	http://www.math.uga.edu/~jtje	index
39	http://www.g4g4.com/mathsocieties1.htm	Math, Puzzles, Games, Mazes, Illusions, Math Puzzles, Math games
40	http://orion.math.iastate.edu/linglong	Ling Long
41	http://www.math.metu.edu.tr/~serge/centers.html	institutions
42	http://www.tundria.com/GSWebDir/SCIEMATH.HTM	MATHEMATICS
43	http://www.math.unl.edu/pi/links	Links - UNL - Department of Mathematics
44	http://arf.math.metu.edu.tr/~dpierce/math.html	Mathematics (David Pierce)
45	http://www.fields.utoronto.ca/resources/links.html	Fields Institute - Visitor Information - Useful Links
46	http://www.math.washington.edu/~palmieri/bookmarks.html	Bookmarks
47	http://faculty.ncwc.edu/clawrence/resources.html	North Carolina Wesleyan College Online Mathematics Resources
48	http://www-math.mit.edu/~kedlaya/mathlinks.html	Kiran Kedlaya's Mathematics Links
49	http://www.math.nwu.edu/~pearsonp/topology.html	Paul Pearson, Northwestern Department of Mathematics
50	http://www.lib.umich.edu/science/math	Shapiro Science Library - Mathematics Subject Guide

According to our original AltaVista search, the MSRI at Berkeley (msri.org) and The Fields Institute (fields.utoronto.ca) have been colinked together 306 times (see Table 4). On Figure 2, the URLs for the two institutes are mapped in close proximity, within the same cluster group. Among the pages producing the 306 colink counts, evidence taken from the first 50 target links shows that The Fields Institute is responsible for at least three site self-colinks:

- 1) <http://www.fields.utoronto.ca/aboutus/IMSI.html>
(The Fields Institute colinks its own home page with the MSRI home page to show that they both belong to the IMSI (International Consortium of Research Institutes in Mathematical Sciences))
- 2) <http://www.fields.utoronto.ca/programs/scientific/04-05/arthurconf/>
(The Fields Institute colinks its home page with MSRI home page to show that both institutes are sponsors of the Conference on automorphic forms and the trace formula, in honour of James Arthur on the occasion of his 60th birthday)

- 3) <http://www.fields.utoronto.ca/resources/links.html>
(The Fields Institute colinks its own home page with the MSRI home page as part of a useful links page directing all visitors to Mathematics Associations/Institutes/Centres Worldwide)

With each of the above links we have a sense of how the Fields Institute views itself relative to the MSRI Berkeley both as another “similar” world class institute (prestige) and as a “co-operational” institute from a social perspective.

A further review of the list of 50 target pages indicates that several directory-type or navigational target pages were retrieved. For instance:

- 1) <http://www.math.upenn.edu/MathSources.html>
University of Pennsylvania, Department of Mathematics’
“Mathematics Sources” Web page
- 2) <http://camel.math.ca/Services/web.html>
Canadian Mathematics Society’s Web page of “Other Mathematics Service Providers”
- 3) <http://www.maths.lth.se/links/mathlink.html>
Lund University, Sweden, directory of “Mathematics on the Web.”

Previous link motivation research pertaining to inlinks to academic Web sites (i.e., Chu, 2003) has shown that directory-type links are very common (i.e., comprising 50%). Since our original WCA list was developed from a directory (i.e., Google: <http://directory.google.com/Top/Science/Math/Research/Institutes/>); we expect other directory-based Web pages to exist for similar navigational purposes. In this study, some of the paired institute pages could be producing higher colink counts because they are listed together in more directories. If this is the case, the WCA is generating a map that is somewhat trivial. We might conclude that it is trivial unless we can find another, more significant reason for the colink configuration. The Web pages that interest us then, are those that have created colinks between the institutes for a purpose other than navigation. A closer look at the retrieved list in AltaVista for msri.org and fields.utoronto.ca reveals that there are such pages:

- 1) <http://www.combinatorics.org/Conferences/>
The Electronic Journal in Combinatorics list all the conferences and meetings held worldwide in Discrete Mathematics
- 2) <http://www.cohomology.com/>
Home Page of Gregory B. Landweber
- 3) <http://www.desargues.univ-lyon1.fr/home/wagner/fowae.html>
Home Page of Frank Olaf Wagner:

At the Electronic Journal in Combinatorics Web page, the MSRI (msri.org) and Fields Institute Web pages (fields.utoronto.ca) receive one colink count because both have been selected as conference venues for researchers in Combinatorics (Discrete Mathematics). The creator of the source page – the colinking page – has shown us that the two institutes share a similar social function. At Gregory B. Landweber’s page, and at Frank Olaf Wagner’s page, the two institutes are colinked again, but this time, for personal reasons. Both scholars have held postdoctoral positions at the MSRI (Berkeley, California) and both have been visitors at different times to The Fields Institute (Toronto, Canada). Compared to the navigational colinks, socially motivated and personally motivated colinks may be less plentiful. If our intent is to determine that they are the primary cause for some of the institutes’ colink proximities, more pages of this type must be uncovered: *Are the institute URLs*

mapped in close proximity colinked more often due to social and personal reasons than those mapped at distance?

Geography seems to have played a significant role in the development of the colink clusters. Note from Figure 2 there is a strong American bias, where several of the Canadian and U.S. institutes are clustered together (e.g. fields.utoronto.ca; pims.math.ca; crm.umontreal.ca; msri.org). In the same cluster, there is also an associated European bias (e.g., newton.cam.ac.uk; mpim-bonn.mpg.de; cirm.univ-mrs.fr; crm.es).

History and/or the rate of Internet technology adoption could be another factor influencing the cluster configuration. Some of the institutes may be receiving more inlinks and colinks because they were the first international conference sites of their kind. Perhaps they have existed longer in the real world and on the Web; thus Web site developers have been inspired to create more inlinks and colinks to their pages.

Web developers' perceptions of the institute environments might also be a cause for some of the colink counts. Certain institutes may be considered more prestigious than others, particularly if they are situated in countries where financial support is generous and research is expected to have a higher international impact. *Are Web page developers colinking specific institute pages more frequently because there is an underlying prestige motivation?* Consider the following page:

- 1) <http://orion.math.uwaterloo.ca/~hwolkowi/intrstsites.html>
Henry Wolkowicz' page of "Interesting Sites" in mathematics:

Although Wolkowicz' page seems to have a navigational purpose, it is also clearly a personal Web page. It is a specific type of personal page (i.e., different from the pages of O. Wagner and G. B. Landweber) designed not only to convey the activities and accomplishments of the creator, but also his personal value judgement. Wolkowicz uses his page to note what he thinks are "interesting" Web page in mathematics; however, underlying this notion of "interest" there might be a prestige motivation. We could investigate this notion further, but as Cronin (1984) suggests "attempts to expose personal motivations are likely to founder, for the simple reason that it is the results, not the process leading up to citation selection [*or link selection*] which authors [*Web site developers*] are only likely to recall if questioned directly on their citation practices [*link practices*]" (p. 57).

3. Interpreting Colink Maps

Unlike the ACA map (Figure 1) the WCA mapping of the international mathematics research institutes (Figure 2) is presented without cluster labels. The proposed strategy for this colink interpretation is exploratory, thus notably incomplete. To complete the interpretation, it is necessary to extract colink lists for every institute pairing, visit and examine the links based on current link motivation theory and carry out a full thematic classification. This type of classification exercise has been done effectively using direct links (e.g., Bar-Ilan, 2005), but with colink data – data that has been amalgamated into a colink network – it is an extensive and highly impractical process.

Bar-Ilan's (2005) significant contribution to link motivation theory has been to examine individual links to academic Web pages, and to show why it is important to take into account both the source and target pages. With direct links, she explains, it is important to remember that they "can be (most often are) created without the knowledge of the author of the target" (p. 978). Consequently, it is usually the "characteristics of the source page that are more influential." (p. 978). Link types can be classified according to the following: a) the relationship between source and target, b) the link tone – i.e., positive; negative; neutral, c) the placement of the link, and d) the link context – i.e., how it is discussed on the source page (pp. 980-981).

Colinks in contrast to direct links, require an extension to Bar-Ilan's (2005) classification, because they incorporate a new dimension for analysis. Figure 4 below illustrates this added dimension. With each and every colinking condition, we want to know why source page A is linked to target page B and why source page A is also linked to target page C. Since B and C are colinked

there is an association, implied by the source page, and although B and C are receiving direct inlinks, they are not necessarily linked to each other (they could be linked, but not necessarily).

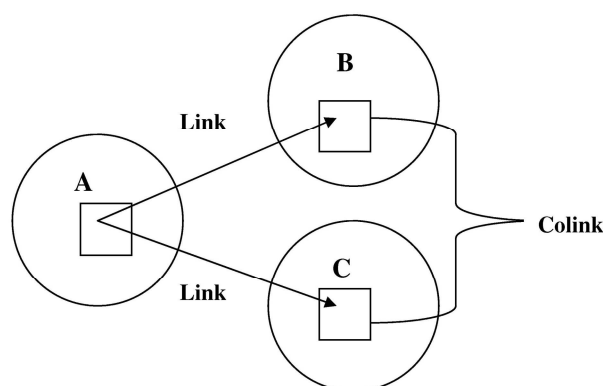


Figure 4. Colink illustration. Source page A links directly to pages B and C. B and C are colinked.

Figure 4 represents a basic example of a colink; thus, when we assemble hundreds of directed link and colink associations together in a full WCA map, we have a mass of data that is too difficult to examine at a microscopic level (see also Harris et al., 2004).

Given the colink map of the mathematics research institutes (Figure 2), we know that the targeted colinked pages possess a “similarity” to one another because the chosen units of analysis are all institute Web pages. Some of the institutes, however, are colinked more often (i.e., in closer proximity) than others. In this case, there is a need for a general classification scheme, highlighting possible reasons for the distances and proximities. With a bottom-up approach to a WCA interpretation, focusing on links, colinks, and motivations for link creations, some reasons may be identified, but without a full examination and ranking of all the links, none of the reasons will be conclusive enough for a definitive interpretation. Our link analysis strategy suggests so far that there could be at least six motivations influencing the institute WCA configuration:

- 1) *Social*: colinks occurring because some of the mathematics institutes have held cooperative research programs and workshops (potentially jointly funded) based on the same subject. Sometimes such colinks are created as site self-colinks.
- 2) *Navigational*: colinks that exist because several different directories on the Web are listing the same groupings of mathematics institute home pages.
- 3) *Personal*: colinks that occur because actual persons (mathematicians) have visited some of the worldwide institutes and present links on their home page to convey the places that they have visited.
- 4) *Geographical*: colinks that have arisen because many Web page developers are aware of the geographical closeness of some of the institutes.
- 5) *Historical*: colinks resulting from the fact that some of the international mathematics institutes have been operational for a lot longer than the others, thus more often recognized on the Web because their home pages have existed for a longer period.
- 6) *Prestige*: colinks that demonstrate an element of prestige associated with certain institute environments because they are well-funded, belong to similar international consortiums and tend to be situated in more affluent countries.

In addition to the bottom-up approach, there is of course another type of examination that may be carried out, and that involves asking some mathematicians or directors of the research institutes to view the map and provide ideas concerning the colink structure. This is an approach commonly used in ACA, where cocitationists often “look for communication ties among authors after clustering them – for example, by interviewing the real people involved” (White, 1990, p. 98).

For this study, a panel of mathematicians (N=3) were recruited to examine the colink map (Figure 2), and what we found with respect to their comments were ideas quite similar to the thematic colink classifications. One of the interviewed mathematicians (i.e., a director from a North American institute) openly stated: “there doesn’t seem to be a really clean interpretation of the clusters. I can make observations though” (personal communication, June 17, 20 and 21, 2005):

A number of institutes operate by choosing a small number of thematic topics each year, and mounting an intensive program of postdoctoral fellowships, conferences, and visitors (both local and international) related to those topics. The idea being to get a critical mass of people together and try to move that field forward in a major way. Many (though not all) of the institutes that operate that way are located in the far right hand cluster. Examples include MSRI, Fields, CRM, and Newton. But that cluster includes other institutes with a different mode of operation (eg IAS).

A different trend is that the clusters seem to segregate institutes by country. For example, all the North American institutes are in the far right hand cluster.

Some of the variations you observe may arise from how seriously the institutes in question use their Web pages. Fields, for example, puts everything up on its Web page, in perpetuity. Other places may do this less.

Other variations will arise because different institutes serve different functions. Some institutes have a very local character – serving the needs only of their country, or even only of their host university. Others (eg Fields, MSRI, Newton) take a truly international point of view. You may also be picking up idiosyncrasies like the fact that CRM and Fields run a prize together, and link to each others' Web sites over that. Or that the BIRS Web site is a subdomain of the PIMS one.

The second mathematician (i.e., a director from one of the European institutes) tended to agree with the first on the subject the map’s overall interpretability: “I’m afraid that there’s not much I can see... many of the linkages don’t seem to be particularly predictable” (personal communication, June 23, 2005):

There is a fair amount of geographical clustering in the results, though less than might be expected. For instance, Eastern European institutes often appear near each other.

The outlying status of AIMS (Aims for Africa) is easily explained... it is not so much an international research institute; more of an educational facility aimed at improving the mathematical skills of Africans at mostly the undergraduate level. Therefore, its activities are of a completely different type to the other institutes.

From the third mathematician (i.e., a frequent visitor to several of the international research institutes), more statements were made regarding geographical closeness, including comments about the clustered institutes’ level of cooperation and mission-type similarities (personal communication June 24, 2005):

Looking at the top left quadrant of your map, it seems that geographical closeness could explain some nearness (Ukraine, Sobolev, Slovak, Czech). [Also], in the top right quadrant there are MSRI, CRM Montreal, Fields. There are surely some

institutionalized links/cooperation between these major North American institutes.

Oberwolfach and the CIRM do have formal agreements between them. More to the point for your analysis perhaps is that they have the same function: weekly meetings throughout the year are organised in the 2 places respectively by the German Math Society and the French Mathematics Society. Also, the CRM Barcelona, the CIM Mexico and IHES seem to have similar functions, as does the IAS Princeton (i.e., special programmes for a year or half year and also long term visitor programmes).

The institutes in the big left region are fairly local, small, not so interesting, but may be inter-supportive.

The major centres: Berkeley, Bonn, IHES, Princeton, Cambridge, Barcelona, Banach Centre, Montreal, Luminy, all attract top mathematicians for one week meetings or longer stays. These are the researchers doing the most important recent research.... The most active (busy, travelling, productive) mathematicians will have regular visits to the major research institutes during their career. So if you look at their personal Web sites you will get a high colink count relating to these research centres.

In sum, what we can take from the interviews is a confirmed recognition of the institute colink map as a valuable observation tool. It may not be possible with Figure 2 to present a set of “clean” interpretive labels, but we can see that the configuration is still informative: it is a valuable Web intelligence resource for mathematicians, particularly institute directors, who might be interested in observing the extent to which the World Wide Web is mirroring their organizational strategies.

3. Conclusion

This study focuses on the relationship between ACA and WCA as “sister” techniques, and shows that WCA is clearly the more complex technique of the two. Author cocitation maps tend to be easier to interpret because they have been created and examined more often than colink maps. ACA is a well-practiced technique and frequent mappings have shown us that the relationship between cocited authors is normally “intellectual,” based on research area subject themes (see White, 1990). When we interpret an ACA structure, we take this “intellectual similarity” at face value; thus we do not examine each and every pair of co-cited author papers. There was one period of time when the purpose and value of ACA was questioned (see Edge, 1977), but again, the principles surrounding this technique are now generally accepted.

With a WCA structure then, should we be expected to visit and classify every type of colink pairing? This is an important question, because it reminds us that colinks on the Web differ significantly from bibliometric cocitations. Colink maps incorporate challenges that we do not see with cocitation maps, both in terms of retrieving link data and selecting links for an interesting analysis. The Web constitutes a much broader thematic context. Web pages lead to various types of mappings, academic or otherwise, thus more colink maps must be created, not only to assess their differences, but to determine also the extent to which link motivation research can help us understand their meaning and research value.

Since link motivation research is still new, there is a lot we can learn from the history of citer motivation research. Leydesdorff (1998) has indicated that a theory of citer motivation must be different at a micro level (i.e., for the individual author and text) than at the aggregate level (i.e., applying to social networks of authors and their systems of concept symbols). The same must be true for a theory of link motivation: what we investigate in terms of inlinks and outlinks at the micro level, may not always be appropriate to an aggregate study of a full link network or colink network. To conclude this ACA-WCA comparative study, a table is presented below with a summary of the two techniques’ similarities and differences.

Table 6. Author Cocitation Analysis (ACA) and Web Colink Analysis (WCA) Comparison.

Author Cocitation Analysis (ACA)	Web Colink Analysis (WCA)
Selection of Data	
Highly cited/cocited authors, selected according to a common, yet diversified subject area, problem area or research field.	Well-linked web pages, selected on the basis of a common theme (e.g., academic web sites)
Data Retrieval	
<p>Boolean AND pairing of authors using Dialog™ citation indexes (e.g., <i>SciSearch@ Social SciSearch</i>®). S CA=Author, A? AND CA=Author, B?</p> <p>Mathematical set theory may be used for complete cocited author retrieval regardless of author rank (Rousseau & Zuccala, 2004).</p> <p>Author homonym pairs from other subjects/fields are possibly retrieved, but occurrence is rare.</p> <p>Data (cocitation counts) are historical in nature – a reflection of the authors’ past work</p> <p>Author self cocitations are normally included in the retrieval so that authors can contribute to their own cocitation profile.</p>	<p>Boolean AND pairing of web pages using AltaVista advanced search window. link:www.domainname.edu AND link:www.domainname.edu</p> <p>Retrieval process includes only pages that link to the home pages of the Web sites.</p> <p>Homonymous pairing of Web pages extremely rare – Web domains are registered.</p> <p>Data (colink counts) are “up to the minute,” but can fluctuate rapidly</p> <p>Researcher should specify why Web site self-colink data is either removed or included during the colink retrieval.</p>
Data Matrices	
<p>Cocitation counts are assembled in an adjacency matrix – all possible pairs can reach a maximum number of $N(N-1)/2$.</p> <p>Pearson’s r or Salton’s Cosine as a similarity measure?</p> <p>Matrix diagonal – Do we scale the data, treat diagonal as missing value, or include self-cocitations?</p>	<p>Colink counts are assembled in an adjacency matrix – all possible pairs can reach a maximum number $N(N-1)/2$.</p> <p>Pearson’s r or Salton’s Cosine as a similarity measure?</p> <p>Matrix diagonal – Do we scale the data, treat diagonal as missing value, or include self-colinks?</p>
Data Mapping and Clustering	
<p>Cocitation mapping produces a scatterplot of author nodes in two or three-dimensional space (i.e., select the best “fit”).</p> <p>Hierarchical-Agglomerative CLUSTER routine in SPSS is used to draw boundary lines around author clusters on the map (cluster cut-off points established through output dendrogram).</p> <p>A core-periphery configuration is evident around the map’s x and y-axes: points representing authors with high similarities are close together in ‘intellectual space,’ while points representing authors with high dissimilarities are farther apart.</p>	<p>Colink mapping produces a scatterplot of Web page nodes (URLs) in two or three-dimensional space (i.e., select the best “fit”).</p> <p>Hierarchical-Agglomerative CLUSTER routine in SPSS is used to draw boundary lines around URL clusters on the map (cluster cut-off points established through output dendrogram).</p> <p>A core-periphery configuration is evident around the map’s x and y-axes: points representing Web pages with high similarities are close together, while points representing Web pages with high dissimilarities are further apart.</p>
Cocited Author/Web Colink Interpretation	
<p>Author maps reveal the “cognitive” or intellectual structure of a field</p> <p>The knowledgeable interpreter of a cocitation map may see much to explicate in the fine structure of author points: for example common nationality, temporal conjunctions, teacher-student relationships, collegial and co-author relationships, or common philosophical orientations (from White, 1990, p. 103).</p>	<p>Colink maps reveal some type of Web structure that is not necessarily intellectual.</p> <p>The knowledgeable interpreter of a colink map may have an intuitive understanding of what the configuration means, but may also need to adopt a bottom-up link analysis approach – i.e., examine extracted colinked pages for predominant relationships, perhaps common geographical connections, mission-based connections, or subject area orientations.</p>

4. References

- Ahlgren, P., Jarneving, B. & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science & Technology*, 54, 550-560.
- Almind, T. C. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'. *Journal of Documentation*, 53, 404-426.
- Bar-Ilan, J. (1999). Search engine results over time: A case study on search engine stability. *Cybermetrics*, 2/3. Retrieved January 10, 2005 from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>.
- Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing & Management*, 41, 973-986.
- Bjorneborn, L. (2003). *Small-world link structures across an academic Web space: A library and information science approach*. Unpublished doctoral dissertation, Royal School of Library and Information Science, Copenhagen, Denmark.
- Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50, 65-82.
- Chu, H. (2003). *Reasons for citation (hyperlinking): what do they imply for Webometric research?* Paper presented at the 9th International Conference on Scientometrics and Informetrics, 25-29 August 2003, Beijing.
- Cronin, B. (1984). *The citation process*. London: Taylor Graham.
- Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science*, 45, 537-538.
- Edge, D. O. (1977). Why I am not a co-citationist. *Society for Social Studies of Science Newsletter*, 2, 13-19.
- Garfield, E. (1965). Can citation indexing be automated? In, M.E. Stevens, V.E. Giuliano, & L.B. Heilprin (Eds.), *Statistical association methods for mechanized documentation, symposium proceedings*, Washington, DC, 1964 (pp. 189-192). Washington, DC: National Bureau of Standards. (NBS Miscellaneous Publication 269).
- Harries, G., Wilkinson, D., Price, L., Fairclough, R. & Thelwall, M. (2004). Hyperlinks as a data source for science mapping. *Journal of Information Science*, 30, 436-447.
- Harter, S. P. (1986). *Online information retrieval: concepts, principles, and techniques*. San Diego, Academic Press.
- Kim, H. J. (2000). Motivations for hyperlinking in scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science*, 51, 887-899.
- Larson, R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of Cyberspace. *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, 71-78. Retrieved January 9, 2005, from <http://sherlock.berkeley.edu/asis96/asis96.html>.
- Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, 43, 5-25.

- Leydesdorff, L. (2005). Similarity measures, author cocitation analysis, and information theory. *Journal of the American Society for Information Science and Technology*, 57(7), 769-772.
- McCain, K. W. (1990). Mapping authors in intellectual space: a technical overview. *Journal of the American society for Information Science*, 41, 433-443.
- Ortega-Priego, J. L., & Aguillo, I. (2005). A Web map of the SCIC research centres: a comparative study of the cosine and Pearson's r . In P. Ingwersen and B. Larsen (Eds.) *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (pp. 197-204). Sweden (Stockholm): Karolinska University Press.
- Perry, C. A., & Rice, R. E. (1998). Scholarly communication in developmental dyslexia: influence of network structure on change in a hybrid problem area. *Journal of the American Society for Information Science*, 49, 151-168.
- Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, 50, 339-344.
- Polanco, X., Boudourides, M. A., Besagni, D. & Roche, I. (2001). Clustering and mapping Web sites for displaying implicit associations and visualising networks. Patras, Greece: University of Patras. Retrieved January 10, 2005 from http://www.math.upatras.gr/~mboudour/articles/web_clustering&mapping.pdf
- Rousseau, R. (1997). Sitations: an exploratory study. *Cybermetrics*, 1. Retrieved January 5, 2005 from <http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>.
- Rousseau, R. (1999). Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics*, 2/3. Retrieved January 5, 2005 from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>.
- Rousseau, R. & Zuccala, A. (2004). A classification of author cocitations: definitions and search strategies. *Journal of the American Society for Information Science and Technology*, 55, 513-529.
- Sandstrom, P. E. (1998). *Information foraging among anthropologists in the invisible college of human behavioral ecology: an author cocitation analysis*. Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.
- Small, H. (1993). Macro-level changes in the structure of cocitation clusters: 1983-1989. *Scientometrics*, 26, 5-20.
- Smith, A. (2004). Web links as analogues of citations. *Information Research*, 9. Retrieved June 6, 2005 from <http://informationr.net/ir/9-4/paper188.html>.
- Snyder, H. & Rosenbaum, H. (1999). Can search engines be used as tools for Web-link analysis? A critical view. *Journal of Documentation*, 55, 375-384.
- Thelwall, M. (2001). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52, 1157-1168.
- Thelwall, M. (2002). The top 100 links to pages on UK university web sites: high inlink counts are not usually directly associated with quality scholarly content. *Journal of Information Science*, 28, 483-491.

- Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8, paper no. 151. Retrieved January 5, 2005 from <http://informationr.net/ir/8-3/paper151.html>.
- Thelwall, M. (2004). *Link analysis: An information science approach*. San Diego: Academic Press.
- Thelwall, M. & Smith, A. (2002). A study of the interlinking between Asia-Pacific university Web sites. *Scientometrics*, 55, 363-376.
- Thelwall, M. & Wilkinson, D. (2004). Finding similar academic web sites with links, bibliometric couplings and colinks. *Information Processing & Management*, 40, 515-526.
- Vaughan, L. & Thelwall, M. (2003). Scholarly use of the web: What are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology*, 54, 29-38.
- Vaughan, L. & You, J. (2005). Mapping business competitive positions using Web co-link analysis. In P. Ingwersen and B. Larsen (Eds.) *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (pp. 534-542). Sweden (Stockholm): Karolinska University Press.
- White, H. D. (1986). Co-cited author retrieval. *Information Technology and Libraries*, 5, 93-99.
- White, H. D. (1990). Author co-citation analysis: Overview and defence. In C. L. Borgman (Ed.) *Scholarly Communication and Bibliometrics* (pp. 84-106). Newbury Park, CA: Sage Publications.
- White, H. D. (2003). Author cocitation analysis and Pearson's r . *Journal of the American Society for Information Science and Technology*, 54, 1250-1259.
- White, H. D., & Griffith, B. C. (1981a). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163-172.
- White, H. D., & Griffith, B. C. (1981b). A cocitation map of authors in judgement and decision research. In B. F. Anderson, (Eds.), *Concepts in judgement and decision research: definitions, sources, interrelationships, comments* (pp. 261-271). New York: Praeger.
- White, H. D. & Griffith, B. C. (1982). Authors as markers of intellectual space: Co-citation in studies of science technology and society. *Journal of Documentation*, 38, 255-272.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author cocitation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49, 327-55.
- Zuccala, A. (2004). *Revisiting the invisible college: a case study of the intellectual structure and social process of Singularity Theory research in mathematics*. Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.