Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 2 (2008) 53–64
Journal of INFORMETRICS
An International Journal
www.elsevier.com/locate/joi

All author cocitation analysis and first author cocitation analysis: A comparative empirical investigation

Sean Eom *

Department of Accounting & MIS, MS 5815, Southeast Missouri State University, Cape Girardeau, MO 63701, USA

Received 30 July 2007; received in revised form 8 September 2007; accepted 10 September 2007

Abstract

The majority of author cocitation analysis (ACA) have relied on the Institute for Scientific Information (ISI) citation databases. ISI convention allows only the retrieval of papers that cite works of which the author is first or sole author. Non-primary authors (authors whose name appear in second or later position) will not be counted when assembling a cocitation frequency matrix. Therefore, this has been a methodological issue in ACA study. This paper empirically examines the impact of the ISI convention on the results of ACA. Previous research has addressed and shed light on some parts of methodological issues, but failed to address issues such as to what extent the use of different approach has resulted in different outcomes in terms of actual intellectual structure of a given academic discipline. Using our data and cociation matrix generation systems, we compare the differences in the process and outcomes of using different cocitation matrices. Our study concludes that all author based ACA is better than first author based ACA to capture all influential researchers in a field. It also identifies more research subspecialties. Finally, all Author based ACA and first author based ACA produce little differences in stress values of MDS outputs.
Published by Elsevier Ltd.

Keywords: Author cocitation analysis; All author ACA; First author ACA

1. Introduction

The majority of author cocitation analysis (ACA) have relied on the Institute for Scientific Information (ISI) citation databases. ISI convention allows only the retrieval of papers citing works of which the author is first or sole author. As McCain (1990, p. 435) states, “Authors who consistently publish as coauthors, in second or later position, will not be visible in the retrievals based on these contributions.”

The purpose of this paper is to empirically examine the impact of the ISI convention of relying only on the name of the first author in assembling the cocitation matrix on the investigation of the intellectual structure of academic disciplines. Virtually all ACA studies using Thomson’s ISI citation indexes only used first author to retrieve the cocitation frequency matrix. Therefore, this has been a methodological issue in ACA study. There are many available full-text research articles and citation indexes that index all authors of cited works on the web (Zhao, 2006). Further, it is possible to use popular software such as Microsoft Excel or Microsoft FoxBASE to develop cocitation generation system (Eom, 2003; McIntire, 2007).

* Tel.: +1 573 651 2615; fax: +1 573 651 2992.
E-mail address: sbeom@semo.edu.

1751-1577/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.joi.2007.09.001

First, literature survey is conducted to review what has been done to deal with this issue. Second, based on the survey of literature, we further argue that previous research have addressed and shed light on some part of methodological issues, but failed to address issues such as to what extent the use of different approaches has resulted in different outcomes in terms of actual intellectual structure of a given academic discipline. Using our data and cociation matrix generation systems, we compare the differences in the process and outcomes of using different cocitaion matrices.

2. Literature survey

There are only a handful of studies (Persson, 2001; Schneider, Larsen, & Ingwersen, 2007; Zhao, 2006) that deal with the impact of different cocitation counting methods on the intellectual structure of a discipline. Prior to reviewing these studies, it is worthwhile to review the method of retrieving first-author cocitation as used in most ACA studies using Dialog Classic systems. The traditional ACA uses retrieval commands below.

? S CR = Simon H? and CR = Keen P?

This command retrieves cocitation counts of all articles in the citation index files when the reference list of a publication includes at minimum one publication with Simon H as first or single author and at minimum one publication with Keen P as first or single author. Consequently, a cocitation count with Simon and Keen ranked in secondary or later positions will never be retrieved.

The Persson Study: Based on 7001 citing papers in the library and information science area in the CD-ROM version of Social Science Citation Index covering 1986–1996, Persson demonstrates that "all author" citation counts should be preferred when visualizing the structure of research fields. "First author" only citation studies distort the picture in terms of the most influential researchers, while the subfield structure tends to be just about the same for both methods. Persson observed that many prominent authors were not present in the intellectual spaces produced by the first author only ACA methodology. Persson disregarded more than 90% of references to papers not indexed by the ISI databases as sources papers. Persson (2001) conclude that first-author based ACA can leave out several influential researchers compared to all author based ACA. Nevertheless, the subspecialties identified by both approaches are almost identical.

The Zhao Study: Zhao compared the results of two different types of cocitation counting: first author only cocitation versus all author cocitation. He distinguished three different definitions of counting cocitations. His approach produced two noticeable results (see Table 1). The first one is, "the picture produced through this simplified all-author cocitation counting contains author groups that are more coherent, and is therefore considerably clearer." The second result is in regard to the number of the research subspecialties. All author based ACA produced fewer research subspecialties than produced through the traditional first-author based ACA analysis (Zhao, 2006).

Zhao used three different ways of counting cocitation counts in his research. They are the traditional first author only method, and inclusive and exclusive all-author cocitation. Zhao compared all three ways of counting cocitations, but used only first five authors rather than all authors and called this approach the simplified approach. Zhao conducted ACA analysis of 312 publications that contain 4578 cited references. The results of Zhao's analysis comparing first-author versus inclusive all-author cocitaion analysis produced the following.

The Schneider et al. study: To extend the Zhao study, Schneider, Larsen, and Ingwersen, 25–27 June, 2007) used the citation index, which is extracted from a corpus of full text XML documents. The XML version of the IEEE Computer Society journal includes 16,819 articles over the period of 1995–2004. The citing 16,819 articles contains a total of 212,657 cited references. Each article on average contains 12.6 references. An arbitrary citation threshold of 75 cited

Table 1
Major results of Zhao's study

	First-author cocitation analysis	Inclusive all-author cocitation analysis
Number of factors	11 factors	4 factors
Number of sub-specialties	Fewer	Larger
Cohesion level of the specialties identified	Lower and less clear picture	High. Producing more coherent, and is therefore considerably clearer picture

authors is used for both all-author and first author ACA studies. Unlike Zhao (2006), who used only first five authors, Schneider et al. used all authors in their in the counting.

Their results indicate that the inclusion of all authors result in producing two-dimensional MDS visualization that better fit the data with lower stress values and all author ACA may lead to stronger concentration in the MDS map.

3. Data and the cocitation generation system

The data for this study were gathered from a total of 692 citing articles in the decision support systems (DSS) area over the past 20 years (1971–1990). These citing articles consist of two categories—270 specific DSS application articles based on the same definition and collection criteria of Eom and Lee (1990) and 422 non-application articles. This study is based on a database file consisting of 259 citing articles from the previous study (Eom, Lee, & Kim, 1993) as well as 11 additional specific DSS articles and 422 non-application articles. These 692 articles were collected from the following three sources: 210 articles over the period of 1975 through 1985 from Elam, Huber, and Hurt (1986); 157 articles over the period of 1969 through 1987 from Sprague and Watson (1989); 203 articles over the period of 1971 through 1988 from Eom and Lee (1990) and 220 additional articles (included to cover the period the other articles did not cover) taken from the same source journals and selected using the same selection criteria used by the three source articles. For specific DSS application articles, Eom and Lee (1990) used the following three criteria: (1) a description of semi- or unstructured decisions; (2) a description of the human–computer interface and the nature of the computer-based support for decision makers' intuitions and judgments and (3) A description of a data-dialogue-model system. For the 422 non-application articles, Elam et al. (1986) chose their list of articles which satisfy one of the following conditions: (1) it must discuss development, implementation, operation, impact, or DSS components. (2) For DSS articles related to contributing disciplines, they must be explicitly related to the development, implementation, operation, use or impact of DSS or DSS components. Sprague and Watson (1989) chose 157 publications because they felt the chosen publications were "the most important, interesting and accessible materials on DSS." Subsequent steps were taken to create a database file that consists of a total of 15,030 *cited* reference records taken from the *citing* articles.

We have developed a cocitation generation system using Fox-base database management systems (Eom, 2003). The system computes author cocitation frequencies between any pair of all (primary and non-primary) authors under study. The system enables the users to overcome a restriction of using ISI databases by giving access to cited coauthors as well as the first authors.

4. Research methodology

The focus of this study is the investigation of the impact of the ISI convention of relying only the name of the first author in assembling the cocitation matrix on the intellectual structure of academic disciplines. Therefore, the author cocitation generator produced two sets of cocitation countsall author and first author only.

4.1. Selection of author and compilation of cocitation frequency matrix

The authors for this study were selected by an objective approach. The majority of ACA research uses a subjective approach, selecting authors based on the researchers' judgment. For example, Schneider et al. chose an arbitrary citation threshold of 75 cited authors. Zhao selects 100 authors from a database based on 312 publications. The authors for this study were selected by a two-stage process.

The first step is the loose screening based on the minimum citation of 60 or more times during the investigation period, based on the prior work of Culnan (1986). Rather than selecting the author set based purely researchers' judgment, the first step of our approach uses citation frequency (60) of each candidate author from 15,030 *cited* references. This stage yields a list of about 100 names, whose citation frequency exceeded the threshold value of 60. For any ACA research using commercial database such as ISI social sciences citation index, it is a very difficult, if not impossible, task due to the time and cost of database search. However, it is a reasonable approach that can be adopted by any ACA researchers using custom databases. There are no predefined strict rules for determining the citation threshold itself.

In the second step, the final author set was determined. While the first step involves the use of citation counts of each author, the second step is based on cocitation counts for further screening. Since the primary focus of ACA is to establish the relationship among selected authors, the second stage is to measure the *connectivity* among the authors selected in the previous step. Due to the possible instability of small cocitation counts, author cocitation analysis researchers introduced several ad hoc criteria for further screening a large pool of candidate authors to finalize a list of authors. The criteria include a *mean cocitation rate* above a certain lower limit per author in each time period (e.g., nine for 10 years of Social Scisearch data), cocitation with at least one-third of the entire author set, or restricting the final author set to the 20% receiving the highest number of citations and cocitations in initial retrieval trials. For further details on several different approaches to compiling a predetermined list of authors, see McCain (1990). However, all these criteria were suggested to be applied to the commercial on-line databases such as SCISEARCH and SOCIAL SCISEARCH.

Our databases are significantly different from those commercial databases in terms of size of records. Besides, the cocitation matrix generation system we developed gives access to cited coauthors as well as first authors. Due to these differences, we could not follow the suggested criteria, e.g. McCain (1990) such as “a mean cocitation rate of ‘*x*’ or more cocitations in each time period.” Rather, we had to invent a new criterion through the method of trial and error. We experimented with the sensitivity of changing the cocitation threshold on the final outcomes (number of meaningful factors to accurately represent DSS research subspecialties). With our databases, we conclude that the number of cocitations of an author with himself/herself can be a better criterion to determine the final author set due to the simplicity. Applying the mean cocitation criteria to our Excel worksheet file (the output from the cocitation matrix generation system) involves too many computations. Whenever we delete/add an author to the final author set, we need to compute the mean cocitation rate of each author again.

Using the cocitation counts of 25 with himself/herself in the period (1969–1990), the initial set of 100 authors was reduced to 67 as the final author set for further analysis. The cocitation count with author himself/herself is a count of the publications in which two or more of the author’s works are jointly cited. Counting the cocitation rate with himself/herself is equivalent to “inclusive all-author cocitaion” counting method (Zhao, 2006). Later, we experimented and lowered the threshold from the 25 cocitation rate to a 20 cocitation rate. Although the number of authors increased from 67 to 80, the number of meaningful factors remained the same, and the major conclusions reported did not change. Also, it is important to point out that cocitation thresholds themselves, as sole connection criteria, are suspect in a highly multidisciplinary area. One should look at the overall connectedness and the focused cocitation counts as well.

4.2. Compilation of cocitation frequency matrix

To overcome a standard problem with the Institute for Scientific Information (ISI) databases which code only the first author of a cited work, a Fox-Base based matrix generation system was developed to compute a cocitation frequency between any pair of authors. The cocitation matrix generation system we developed gives access to cited coauthors as well as first authors.

In order to investigate the impact of the ISI convention of relying only the name of the first author in assembling the cocitation matrix on the identification of the intellectual structure of DSS field, we have assembled two cocitation matrices using two different cocitation counting methods (first author only counting and inclusive all author counting). The cocitation matrix using the inclusive all author counting method consists of 67 author variables and 67 observations. The first author only counting method produced a matrix of 47 variables and 47 observations. The different number of authors, 67 versus 47, under each counting method must be interpreted as an important factor that will produce different intellectual structures of an academic discipline. Recall that Zhao selected the same number (100 authors) of authors under the two different cocitation counting methods. This is an important issue for further clarification. As said earlier, we used two different stages of the author selection process to be included in the final author set. The first step is based on citation frequencies. The second step is cocitation frequencies. Applying the same threshold of citation and cocitation frequencies to the two cocitation frequency matrices generated by the first author and all author counting methods should produce different numbers of the final author set. This is because the non-primary (first) authors do not receive any credit at all. Consequently, non-primary authors cocitation counts substantially decrease under the first author only counting method.

There are several ad hoc criteria for those who are using commercial databases, as discussed. Although ACA researchers using other custom databases cannot blindly apply these criteria, it is still important to consider the possible instability of small cocitation counts. It means that the first author only cocitation counting method should produce a smaller number of the final author set.

4.3. Normalization of raw cocitation frequency data

Whether cocitation data should be normalized or not is an important issue in cocitation analysis. Waltman and van Eck (2007) recently refuted the assertion by Leydesdorff and Vaughan (2006) and argued that normalizing the cocitation matrix has been widely adopted and it is nothing wrong with this normalization practice. Raw cocitation frequency data is normalized when using the factor analysis procedure of the SAS system. Furthermore, the cluster and multidimensional scaling (MDS) procedures of the SAS system use distance matrices. The distance matrix is another normalized data from the raw cocitation frequency matrix. The PROC DISTANCE procedure of the SAS system normalizes the cocitation matrix into a matrix of correlation transformed to Euclidean distance as square root of (1 − correlation). A transposed cocitation matrix is processed by the SAS system's factor procedure. The factor procedure converts the matrix into a Pearson correlation matrix without explicit command 'PROC CORR' unless the diagonal value of the matrix is treated as missing values. If that is the case, the factor procedure of the SAS takes the output from the correlations procedure (PROC CORR), Pearson correlation matrix, as the input to the factor procedure.

```
 * create correlation matrix;
proc corr data=sample outp=test;
run;

* input correlation matrix into FACTOR;
proc factor data=test;
run;
```

4.4. Results from the factor analysis

The two cocitation matrices of 57 authors and 47 authors are processed by the SAS factor procedures. Inclusive all author cociation matrix is analyzed by the principal component analysis (PROMAX rotation) with the latent root criterion (eigenvalue 1 criterion) to obtain the initial solution of 9 factors (see Table 2). The detailed discussion of these results can be found in Eom and Farris (1996). Table 3 shows the factor analysis results of first author only cocitation matrix.

The study of Schneider et al. (2007) produced a 29-factor model for the conventional approach to first author cocitation analysis, which explains 59% of the variance; a 25-factor model for the conventional approach to all author cocitation analysis, which explains 63% of the variance. The summary of the two previous studies as shown in Table 4 indicated that the first author cocitation study produced more factors while explaining less percentage of variances. This result is most contrasting when compared with this study of Eom. Contrary to the conventional wisdom (increasing number of factors increases the percentage of variance explained), increasing number of factors decreases the percentage of variance explained. The message is clear. While both previous studies (Zhao and Schneider et al.) kept the same number of authors in the cocitation matrices, the instability of small cocitation counts in the cocitation matrices increased.

Of these three studies, the most important implication is that first author only cocitation analysis undeniably decrease the percentage of variance explained. By omitting non-primary authors contributions, we may misrepresent the intellectual structure of an academic discipline.

Secondly, by keeping the same number of authors the researchers are producing the results of increasing number of factors that can potentially poorly represent the research subspecialties.

Finally, the result of Eom's study clearly shows that not counting non-primary authors contribution in ACA analysis may results in loss of several meaningful factors.

4.5. Multidimensional scaling analysis

A study by Zhao claimed that "the picture produced through this simplified all-author cocitation counting contains author groups that are more coherent, and is therefore considerably clearer." Schneider et al. (2007) made a similar claim that "the all-author ACA maps result in stronger concentration of the co-cited authors into clusters, regardless of the approach to matrix generation."

Table 5 shows stress 1 values of all studies. The two studies of Eom and Schneider produced very similar results in term of badness of fit statistics.

Table 2
Factor structure correlations (1970–1990) using inclusive all author cocitation counting method

Factor 1 foundations		Factor 2 GSS		Factor 3 model management	
KEEN	0.95993	GALLUPE	0.95129	WHINSTON	0.94146
SCOTTMORTON	0.95944	HILTZ	0.91792	BONCZEK	0.94119
ALTER	0.92275	GRAY	0.91571	HOLSAPPLE	0.93471
CARLSON	0.89093	BUI	0.91336	BLANNING	0.91452
SPRAGUE	0.88596	TUROFF	0.9113	ELAM	0.90118
GORRY	0.83964	KRAEMER	0.90857	HENDERSON	0.86374
GINZBERG	0.83205	KINGJL	0.89616	STOHR	0.8601
LITTLE	0.82385	APPLEGATE	0.89411	DOLK	0.82831
SIMON	0.8161	DESANCTIS	0.86535	GEOFFRION	0.78473
BENNETT	0.80797	NUNAMAKER	0.85517	DAVISR	0.75146
ANTHONY	0.80608	GEORGE	0.80845	KONSYNSKI	0.72498
KINGW	0.79081	HUBER	0.77487	SPRAGUE	0.68514
WATSONH	0.78638	JARKE	0.7159	SHORTLIFFE	0.67171
WAGNERGR	0.77806	DELBECQ	0.68703	CARLSON	0.65223
MEADOR	0.77143	VANDEVEN	0.68354	NAYLOR	0.64614
DAVISGB	0.75277	KONSYNSKI	0.66861	WATSONH	0.61578
ROCKART	0.73428	JARVENPAA	0.57399	BENNETT	0.57433
NAYLOR	0.71531	JELASSI	0.51461	COURTNEY	0.56916
COURTNEY	0.649	DICKSON	0.46365	MEADOR	0.5519
ROBEY	0.62416			SIMON	0.52894
ACKOFF	0.58481			KEEN	0.52188
ZMUD	0.57803			JARKE	0.52172
STOHR	0.56825			SCOTTMORTON	0.49578
LUCAS	0.5586			ALTER	0.47525
MINTZBERG	0.55572			JELASSI	0.45785
HENDERSON	0.54863			GINZBERG	0.41872
SANDERS	0.54786			HUBER	0.41159
CHERVANY	0.53879				
SHORTLIFFE	0.52861				
DAVISR	0.50259				
WHINSTON	0.49768				
HOLSAPPLE	0.4857				
MASON	0.4812				
BENBASAT	0.48099				
NEWELL	0.48089				
BONCZEK	0.47731				
MITROFF	0.43538				
HUBER	0.43043				
BLANNING	0.41798				
Variance	19.28		12.63		14.94
Percentage of variance	28.78		18.85		22.30

Factor 4 user interface		Factor 5 organizational science		Factor 6 MCDM	
BENBASAT	0.95209	MARCH	0.8423	KEENEY	0.9452
DEXTER	0.92254	SIMON	0.83475	RAIFFA	0.92685
LUSK	0.90435	NEWELL	0.81959	GEOFFRION	0.66104
LUCAS	0.90213	MINTZBERG	0.74919	JELASSI	0.60025
DICKSON	0.88478	TVERSKY	0.74032	JARKE	0.52554
ZMUD	0.86408	ACKOFF	0.67446	STOHR	0.43314
IVES	0.85857	ANTHONY	0.55723		
CHERVANY	0.84881	KEEN	0.53412		
ROBEY	0.72918	SCOTTMORTON	0.53412		
JARVENPAA	0.67287	DAVISGB	0.50856		
MASON	0.61158	GORRY	0.49208		
MITROFF	0.60447	COURTNEY	0.49183		
KINGW	0.5693	DAVISR	0.48454		
SIMON	0.55545	MITROFF	0.4839		

Table 2 (*Continued*)

Factor 4 user interface		Factor 5 organizational science		Factor 6 MCDM
DAVISGB	0.55099	SHORTLIFFE	0.48016	
KEEN	0.53746	MASON	0.46292	
COURTNEY	0.52986	HUBER	0.46097	
DESANCTIS	0.51727	SPRAGUE	0.4591	
SCOTTMORTON	0.515	MEADOR	0.44851	
MINTZBERG	0.50778	SANDERS	0.44539	
SANDERS	0.48077	CARLSON	0.43658	
GINZBERG	0.46672	BENBASAT	0.43317	
HUBER	0.44862	ZMUD	0.42381	
NEWELL	0.43114	LUCAS	0.42011	
GORRY	0.42748	ROBEY	0.40571	
ACKOFF	0.40943	DICKSON	0.40095	
Variance	10.96		4.42	2.77
Percentage of variance	16.36		6.60	4.13

Factor 7 strategic management		Factor 8 implementation		Factor 9 group decision making	
MITROFF	0.66403	SANDERS	0.68	VANDEVEN	0.44691
MASON	0.64888	COURTNEY	0.63997	DELBECQ	0.43894
ACKOFF	0.44547	HENDERSON	0.51111	NUNAMAKER	0.4124
		HUBER	0.45316		
		ZMUD	0.43705		
		ROBEY	0.41624		
		KONSYNSKI	0.41312		
Variance	2.77		5.07		4.42
Percentage of variance	4.13		7.57		6.60

The purposes of MDS are to help researchers identify the "hidden structures" in the data and visualize relationships among/within the hidden structures to give clearer explanations of these relationships to others (Hair, Anderson, & Tatham, 1987; Kruskal & Wish, 1990). Three SAS procedures (MDS, PLOT and G3D) are necessary to convert the author cocitation frequency matrix to two or more dimensional pictures of data. The distance matrix produced earlier by using the PROC distance procedure should be converted to a coordinate matrix. The coordinate matrix is used to produce two-dimensional plots and annotated three-dimensional scatter diagrams. The PLOT and G3D procedures process the coordinate matrix to visualize the similarity and dissimilarity within each group of an academic discipline as well as among the various subspecialties within an academic discipline. In ACA study, 3D scatter plots without labels on data points provide little information for the ACA researchers. The annotate facility in the SAS system produces figures with the name of the author on each data point. The PROC MDS procedure includes many of the features of the ALSCAL procedure (Young, Lewyckyj, & Takane, 1986).

The three-dimensional MDS map as shown in Fig. 3 (the badness of fit value is 0.17) enabled us to map the overall relationships between DSS subspecialties and contributing disciplines, as well as the interrelationships within both the DSS subspecialties and the contributing disciplines to DSS. Multidimensional scaling maps show the big picture of inter-cluster relationships. The placement of authors on the center of the MDS map means that those authors are linked with a substantial portion of the author set, with relatively high correlations (e.g., founding fathers of DSS–Keen, Scott Morton, Alter, Sprague and Carlson). Placement near the periphery represents a more focused linkage. This is illustrated by the model management researchers (Blanning, Bonczek, Whinston and Holsapple) located in the upper-left hand area and GDSS researchers (George, Bui, Gallupe, etc.) located in the upper-right hand area. The location of the organizational/cognitive scientists, especially Tversky, Simon, Newell and Mintzberg) in the center of the MDS map seems to indicate that organizational science and cognitive science are major contributing disciplines in that it connects with many other DSS research subspecialties. Figs. 1 and 2 present two-dimensional MDS map, which is the projection of the first and second planes presented in the three-dimensional MDS map in Figs. 3 and 4.

Table 3
Factor structure correlations using first author cocitation method

Factor 1 foundations		Factor 2 GSS		Factor 3 user interface	
ALTER	0.91687	GALLUPE	0.95402	BENBASAT	0.94049
KEEN	0.89528	BUI	0.93725	DICKSON	0.93806
SCOTTMORTON	0.86716	KRAEMER	0.92921	LUSK	0.89513
LITTLE	0.8618	GRAY	0.9289	IVES	0.88806
CARLSON	0.85774	TUROFF	0.92587	LUCAS	0.88661
GORRY	0.8474	NUNAMAKER	0.8952	ZMUD	0.83389
SPRAGUE	0.8376	HILTZ	0.88753	JARVENPAA	0.67097
BENNETT	0.80896	DESANCTIS	0.88339	MASON	0.63275
SIMON	0.77839	APPLEGATE	0.86568	ROBEY	0.59051
ANTHONY	0.76278	HUBER	0.78728	DESANCTIS	0.54792
GINZBERG	0.72609	JARVENPAA	0.64264	KEEN	0.47775
DAVISG	0.69381			SIMON	0.44752
MEADOR	0.68165			HUBER	0.4305
NAYLOR	0.67597			MINTZBERG	0.4237
KINGW	0.66533			SANDERS	0.42011
BONCZEK	0.59856			KINGW	0.41864
ROCKART	0.58887			HENDERSONJ	0.40287
ACKOFF	0.52042				
DAVISR	0.5129				
WAGNERG	0.49836				
ROBEY	0.47872				
MASON	0.46959				
MINTZBERG	0.44699				
BLANNING	0.44565				
LUCAS	0.44516				
HENDERSONJ	0.4232				
SANDERS	0.41644				
Variance	13.543		9.068		8.711
Percentage of variance	28.81		19.29		18.53

Factor 4 model management		Factor 5 implementation		Factor 6 organizational science		Factor 7 unnamed	
ELAM	0.95338	KEEN	0.82242	NEWELL	0.84263	DAVISR	0.48324
BLANNING	0.93628	SPRAGUE	0.81564	SIMON	0.83724	MEADOR	0.42818
KONSYNSKI	0.88701	SANDERS	0.79898	ACKOFF	0.77974	MASON	0.47874
BONCZEK	0.8851	ROBEY	0.77875	MINTZBERG	0.75238		
DOLK	0.85898	HENDERSONJ	0.77072	ANTHONY	0.63564		
SPRAGUE	0.5662	KINGW	0.76055	DAVISG	0.58466		
DAVISR	0.54314	GINZBERG	0.74832	GORRY	0.57882		
NAYLOR	0.51744	ALTER	0.72773	MASON	0.56093		
SIMON	0.41588	ZMUD	0.71422	KEEN	0.52201		
		WAGNERG	0.67873	ZMUD	0.47594		
		MASON	0.67593	SPRAGUE	0.46798		
		MINTZBERG	0.67128	LUCAS	0.4464		
		ROCKART	0.65174	ROCKART	0.43152		
		MEADOR	0.61265	HUBER	0.42293		
		SIMON	0.59713	BENBASAT	0.4187		
		LUCAS	0.57136	SANDERS	0.4035		
		BENBASAT	0.55944				
		GORRY	0.55607				
		DAVISG	0.55355				
		NAYLOR	0.5264				
		HUBER	0.48529				
		ANTHONY	0.47211				
		BONCZEK	0.46762				
		ACKOFF	0.46661				
		BENNETT	0.45853				
		CARLSON	0.44332				
		DICKSON	0.44202				
Variance	6.813		11.879		7.614		1.395
Percentage of variance	14.50		17.73		11.36		2.08

Table 4
Comparison of the results of this study and previous two studies

Researchers	All author ACA	First author ACA
Zhao		
Number of factors	5 factors	11 factors
Percentage of variance explained	97	96
Number of authors	100	100
Database (citing and cited) size	312 publications/4578 citations	Same as left
Schneider et al.		
Number of factors	25 factors	29 factors
Percentage of variance explained	63	59
Number of authors	75	75
Database (citing and cited) size	16,819 articles/212,656 citations	Same as left
Eom		
Number of factors	9 factors	7 factors
Percentage of variance explained	84.82	81.94
Number of authors	57	47
Database (citing and cited) size	692 articles/15,039 citations	Same as left

5. Discussion and conclusions

The present study reviews several previous studies that compare the results of ACA using the traditional first author only and all author method. Three conclusions can be reached based on our study. Nonetheless, we realize that the current study deals with a discipline different that of previous studies. Thus, the different results obtained may not all be attributed to the way the cocitation data were collected.

5.1. All author based ACA is better to capture all influential researchers in a field than first author based ACA

There are only three studies that investigate the impact of different cocitation counting methods on the intellectual structure of an academic discipline. However, all three studies suffer from flaws in either data or methodology. A critical deficiency of Persson’s study is that more than 90% of references to papers not indexed by the ISI databases were left out of consideration. Despite the deficiency of the data, Persson concludes that all author based ACA is better to capture all influential researchers in a field. Eom’s study supports Persson’s conclusion. All author based ACA identified 67 authors while first author based ACA captures only 46 authors.

5.2. All author based ACA identifies more subspecialties

We have shown that the ISI convention of relying on only the name of the first author in assembling the cocitation matrix may often fail to identify all possible underlying factors. This is the most significant conclusion. The most

Table 5
Stress 1 values for MDS configurations

Study	Configuration	First author only method	All author met3hod
Schneider et al.	Two dimensions	0.251	0.219
	Three dimensions	0.186	0.181
Zhao	Two dimensions		
	Three dimensions		
Eom	Two dimensions	0.2530	0.2532
	Three dimensions	0.1768	0.1733
	Four dimensions	0.1278	0.1282
	Five dimensions	0.1010	0.1016
	Six dimensions	0.0811	0.0819

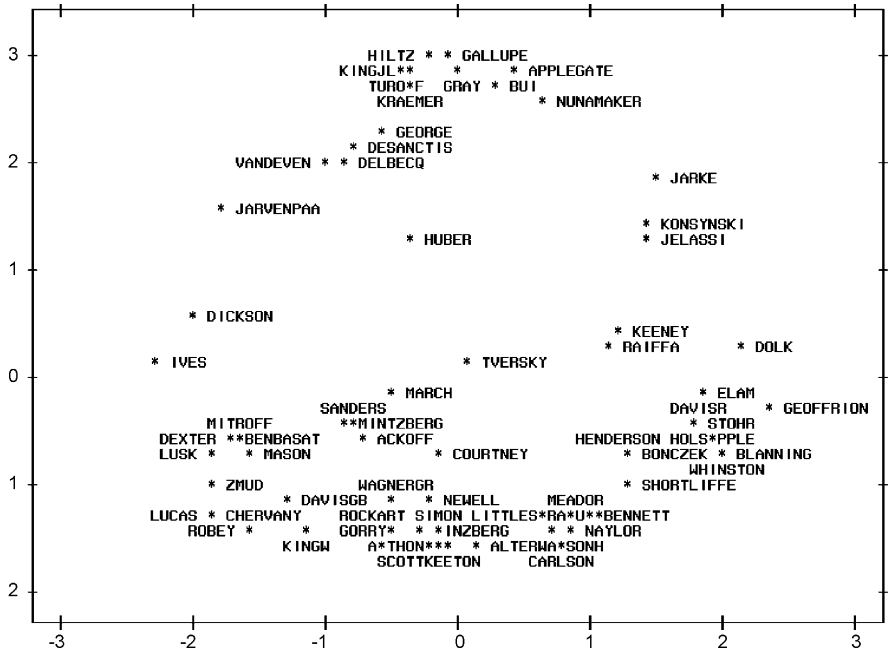

Fig. 1. All author two-dimensional map (67 variables).

important implication of these three studies is that first author only cocitation analysis undeniably decreases the percentage of variance explained. By omitting non-primary authors contributions, we may possibly misrepresent the intellectual structure of an academic discipline. Disregarding non-primary authors contribution in ACA analysis results in the loss of several meaningful factors. When comparing the two Tables 2 and 3, counting only the first author produced only six factors. The three factors (MCDM, strategic management and group decision making) were not visible.

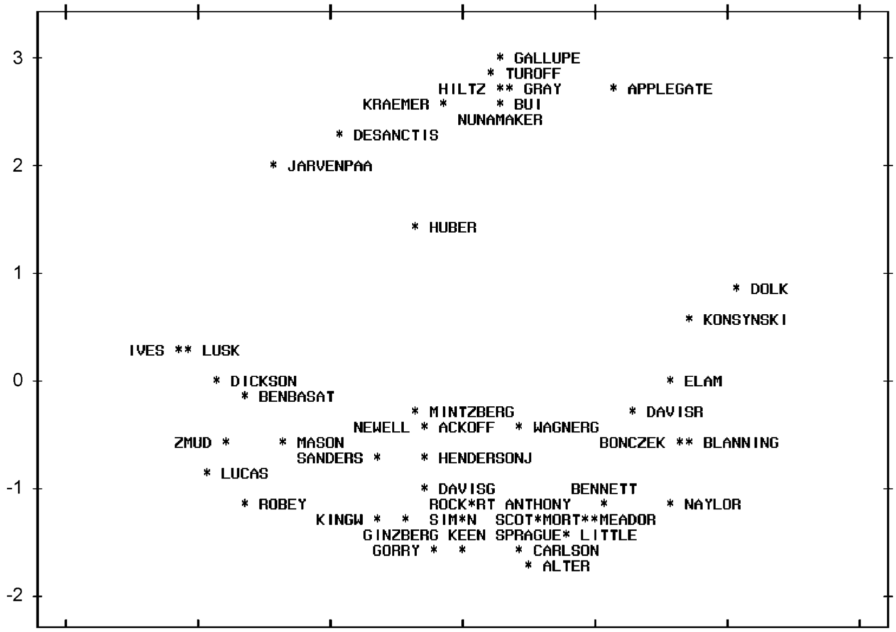

Fig. 2. First author only two-dimensional map (47 variables).

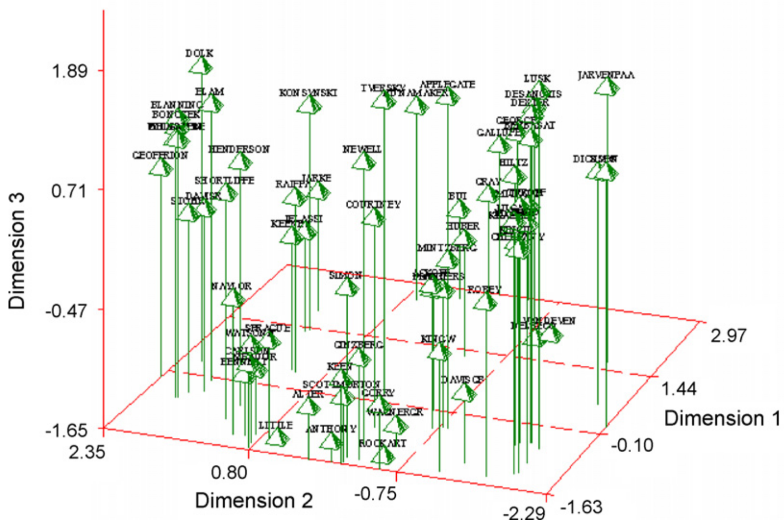

Fig. 3. All author 3D map (67 variables).

This study, therefore, disconfirms the results of the two previous studies by Zhao and Persson in regard to the number of the research subspecialties. Persson claims that the subspecialties identified are almost identical regardless of cocitation counting methods employed. On the other hand, Zhao concludes that all author based ACA produced fewer research subspecialties than produced through the traditional first-author based ACA analysis. The important question when interpreting the result of any multivariate statistical studies including ACA studies is the meaning of each factor. Determining the optimal number of factors in ACA analysis critically depends on the interpretability of each factor, not the eigenvalue criterion, not the scree test result. The interpretability of each factor should be the most important factor to determine the final factor solution.

The first author only ACA result of Zhao's study vividly shows that many of the factors that appeared in the final solution (Table 4 of Zhao's study) may possibly have a problem of interpretability. According to McCain (1990, p. 440), "Only authors with loadings greater than ±0.7 are likely to be useful in interpreting the factor, and only loadings above ±0.4 or ±0.5 are likely to be reported." Applying this criterion to Table 4, many factors in his Table 4 may not be good candidates for the final solution.

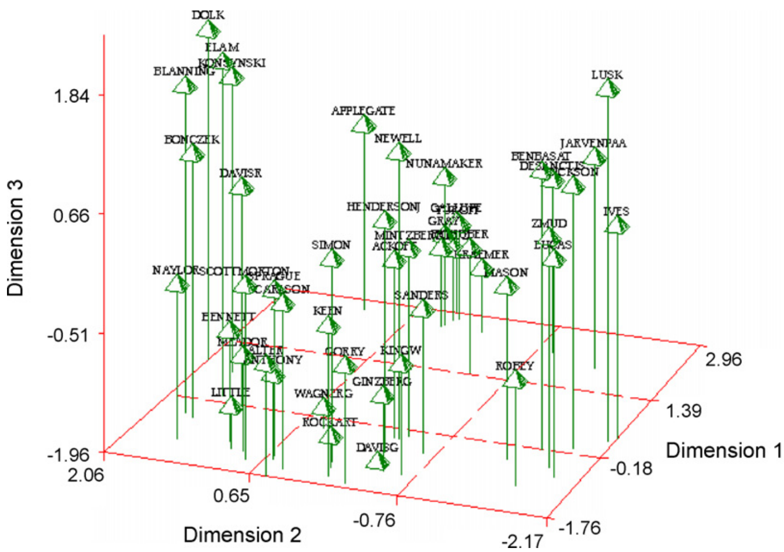

Fig. 4. First author only 3D map (47 variables).

As said earlier, counting only the first author should decrease the cocitation frequencies in a given bibliographic database. If so, the final author set for ACA study should be decreased in number if there is a selection criteria to be applied whether it is a *mean cocitation rate or* cocitation with at least one-third of the entire author set, or restricting the final author set to the 20% receiving the highest number of citations, or any other criterion.

5.3. All author based ACA and first author based ACA produce little differences in stress values

The results of MDS analysis in this study does not seem to support the major conclusion of Zhao: “the picture produced through this simplified all-author cocitation counting contains author groups that are more coherent, and is therefore considerably clearer.” When comparing Zhao’s (Figs. 1 and 2), we agree his conclusion that all-author cocitation based ACA produced a two-dimensional map showing a more focused linkage at the right top section as well as at the bottom middle section.

Visual inspection of the two MDS two-dimensional maps of Zhao (Figs. 1 and 2) does not show such changing configuration on the map. Stress 1 values of the configurations (Table 5) show the badness of fit. The badness of fit is equivalent to Young’s s-stress formula 1. As shown in Table 5, all author based ACA and first author ACA produce little differences in stress values. For example, our study results show that all author based ACA produced larger stress 1 values, which means a poor fit. The real question is how statistically significant the difference is between the two. In the case of two-dimensional configuration, the difference in stress value is 0.0002. We do not believe that this is a significant difference that can reach different conclusion.

References

Culnan, M. J. (1986). The intellectual development of management information systems, 1972–1982: A co-citation analysis. *Management Science*, *32*(2), 156–172.

Elam, J. J., Huber, G. P., & Hurt, M. E. (1986). An examination of the DSS literature (1975–1985). In E. R. McLean & H. G. Sol (Eds.), *Decision support systems: A decade in perspective* (pp. 239–251). Amsterdam, North-Holland: Elsevier Science.

Eom, S. B. (2003). *Author cocitation analysis using custom bibliographic databases—An introduction to the SAS approach*. Lewiston, New York: The Edwin Mellen Press.

Eom, S. B., & Farris, R. (1996). The contributions of organizational science to the development of decision support systems research subspecialties. *Journal of the American Society for Information Science*, *47*(12), 941–952.

Eom, H. B., & Lee, S. M. (1990). A survey of decision support system applications (1971 to April 1988). *Interfaces*, *20*(3), 65–79.

Eom, S. B., Lee, S. M., & Kim, J. K. (1993). The intellectual structure of decision support systems (1971–1989). *Decision Support Systems*, *10*(1), 19–35.

Hair, J. F., Jr., Anderson, R. E., & Tatham, R. L. (1987). *Multivariate data analysis with readings* (2nd ed.). New York: Macmillan Publishing Company.

Kruskal, J. B., & Wish, M. (1990). *Multidimensional scaling*. Newbury Park, California: Sage Publications.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science and Technology*, *57*(12), 1616–1628.

McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, *41*(6), 433–443.

McIntire, J. S. (2007). *The clothing and textile research base: An author cocitation study*. Unpublished master’s thesis. Columbia, Missouri: University of Missouri.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, *50*(2), 339–344.

Schneider, J. W., Larsen, B., & Ingwersen, P. (2007). Comparative study between first and all-author co-citation analysis based on citation indexes generated from XML data. In *Paper presented at the 11th international conference on scientometrics and informetrics*.

Sprague, R. H., Jr., & Watson, H. J. (1989). DSS bibliography. In R. H. Sprague Jr. & H. J. Watson (Eds.), *Decision support systems: Putting theory into practice* (2nd ed., pp. 403–413). Englewood Cliffs, NJ: Prentice Hall.

Waltman, L., & van Eck, N. J. (2007). Some comments on the question whether co-occurrence data should be normalized. *Journal of the American Society for Information Science and Technology*, *58*(11), 1701–1703.

Young, F. W., Lewyckyj, R., & Takane, Y. (1986). *The ALSCAL procedure*. Cary, NC: SAS Institute.

Zhao, D. (2006). Towards all-author co-citation analysis. *Information Processing & Management*, *42*(6), 1578–1591.