Survey of Collaboration for Digital Libraries among Scientists: a bibliometric approach

A. Hossein Farajpahlou[1]

1 Department of Library and Information Science

Shahid Chamran University, Ahvaz, Iran.

Farajpahlou_a@scu.ac.ir

Abstract

Many aspects of digital libraries have been considered and examined so far by many researchers and authors. How much these efforts have been collaborative within the last few years by those who have been actively involved in recording of studies done on theoretical and practical aspects of digital libraries, was the main point of concern for this researcher. To achieve this, the EBSCOs Academic Search FullTEXT Premier database and Science Direct online database were consulted for online electronic articles on digital libraries. The publication period was limited to 5 years, i.e., from January 2000 to January 2005. Some 130 articles in EBSCO database, and 115 articles were returned in Science Direct database. In the course of examination, Collaboration Coefficient equation was used as was introduced by Ajiferuke, et al, [1]. Calculation of the CC values revealed that collaboration in publication of digital library records has been a growing trend in the last five years, though not very much remarkably. 109 subjects were identified in the articles that were discussed in relation to digital libraries. Among these, the highest collaboration rates were identified to be associated with issues such as technicalities, application, experiments, performance, and, information processing and retrieval in digital libraries.

1. Introduction

The term digital libraries and its current concept seem have come into our literature since 1990s. Searches into the relevant literature returns publications that go back to the first half of the decade 1990s. By the turn of the 20th century, the concept and idea of *digital library* was almost common place worldwide, especially in the western world. Many ventures from simple digital collections of electronic lists of journals holdings of an institution, to sophisticated systems and collections of ebooks provided for distance access and use, are nowadays within reach of those who are connected to the internet, or can access electronic library collections whether through OPACs or CD-ROM holdings. On the other hand, it is common sense, as well as experienced nowadays that collaborative scientific activities are very effective in advancing the related areas of knowledge and their related technicalities. Development of multi-disciplinary branches of knowledge since the late 20th Century, not only has resulted in coming new fields of study and new experts into being, but also, has required cooperation and collaboration of different experts to communicate on a multi-disciplinary issue. Therefore, collaboration looks to be the specific characteristic of the new Century in academic and research world. It is now common sense that successful scientific probe to present day issues requires "collaboration" among several experts each of whom deals with a specific aspect in that issue. This is expected to be true with "Digital Libraries" as a new multi-disciplinary concept. Therefore, to some extent, degree of success in this area depends on the degree people who are involved in related aspects and issues collaborate. Based on this logic, the present researcher applied content analysis of relevant

articles on digital libraries to measure the degree of collaboration among authors who wrote on this issue.

2. Method

Two full-text databases of EBSCO's Academic Search FullTEXT Premier and Science Direct were searched for articles written on "Digital Libraries" within the last 5 years, ie, from January 2000 to December 2004. Some 227 articles were returned in EBSCO database, and 118 articles in Science Direct database. The returned articles were then examined for identification of their authors as well as their key subjects which were identifiable through their titles. The data extracted from the count of authors and subjects were then organized and ordered in tables to be prepared for calculation of the Collaboration Coefficients. Collaboration Coefficient (CC) as was introduced by Ajiferuke, et al, denotes the degree of collaboration among a group of authors in a discipline during a certain period of time[1]. CC which is a function of the number of authors for articles, as well as the total number of the collected articles, could be calculated according to the following equation:

$$CC = 1 - \sum_{j=1}^{k} [(1/j)\, f_j / N]$$

Where:

F_j = the number of *j* authored research papers in a discipline in a certain period of time

N = total number of papers published in a discipline during a certain period of time

K = the greatest number of authors per paper in a discipline

J = number of authors

Equation 1, as Ajiferuke [1] indicates, and Havemann [2, 3] expresses, "is a simple, suggestive, and stable indicator for the degree of collaboration and has a direct meaning for the single scientist, too." Prior to this, two other methods were used to be applied, namely, Collaborative Index (CI) which calculates the mean number of authors per paper, and Degree of Collaboration (DC) that calculates the strength of collaboration. As Ajiferuke, et al [1] clearly state, this single equation "incorporates some of the merits of both" methods and excludes their insufficiencies. CI, although differentiates among levels of authorship and is very easy to calculate, but: 1) it is not easily interpretable as a degree for it has no upper limit. Neither it lies between 0 and 1, nor it could be expressed in terms of percentage; 2) it gives non-zero weight to single-authored papers which involve no collaboration. DC, does not have all the CI insufficiencies. It is easy to calculate, it is easily interpretable in terms of percentage, it gives zero weight to single-authored papers, however, it does not differentiate among levels of multiple authorship. Therefore, CC seems to be the most suitable and reliable method of all presently existing for calculation of "collaboration" in a discipline in a certain period of time.

The articles resulted from search in the mentioned international databases, were examined one by one, and the number of authors and subject keywords were recorded. The sub-totals were then calculated for each year. The maximum number of authors did not exceed 7, and the minimum was 1 author per article. The only exception was one article with more than 10 authors which appeared in the Science Direct collection of records for the year 2004. The resulting figures are presented in tables 1 and 2, for EBSCO and Science Direct, respectively.

Table 1: Distribution of articles according to number of authors as derived from Science Direct database

Year	Number of authors per article									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**11**	**Total**	**CC**
2000	14	4	7	2	1	0	0	0	28	0.3440
2001	6	4	1	4	1	0	0	0	16	0.4042
2002	9	8	4	6	0	1	0	0	28	0.4286
2003	14	6	2	1	0	1	0	0	24	0.2465
2004	10	2	3	2	0	1	1	1	20	0.3437
Total	**53**	**24**	**17**	**15**	**2**	**3**	**1**	**1**	**116**	**0.349**

Table 2: Distribution of articles according number of authors as derived from EBSCO database

Year	Number of authors per article								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Total**	**CC**
2000	9	4	7	1	1	0	0	22	0.3736
2001	6	4	1	3	1	0	0	15	0.3812
2002	8	8	4	5	0	0	0	25	0.4127
2003	12	6	2	1	0	1	0	22	0.2691
2004	8	2	3	2	0	0	2	17	0.3656
Total	**43**	**24**	**17**	**12**	**2**	**1**	**2**	**101**	**0.3614**

As it is clear from tables 1 and 2, although the yearly CC values are not quite similar for both databases, however, there appears to be a similarity of fluctuations of CC values in both databases. That is, the values rise up in 2001 and follow a steady progress up to year 2002. Then, there is a fall, and a steady growth again, afterwards. This situation can be clearly observed in figures 1 and 2.

Fig. 1: Fluctuations of the CC values in Science Direct articles between 2000 and 2004

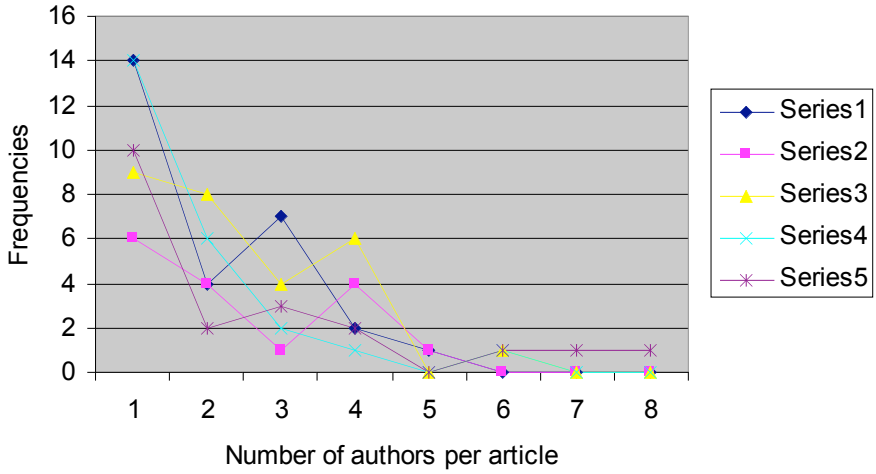
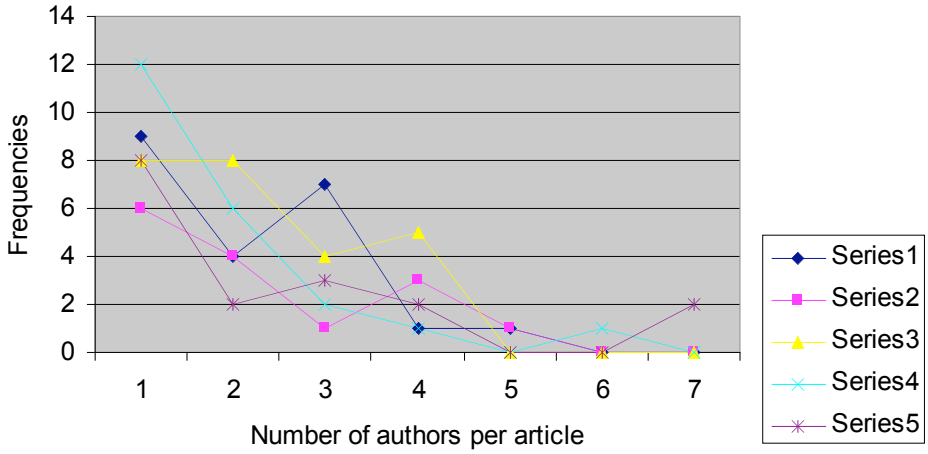


Fig. 2: Fluctuations of CC values for the period 2000 to 2004



In figures 1 and 2, series 1 through 5 represent years 2000 through 2004, respectively.

Another main issue is that the CC values normally rest between 0 and 1. The more it is closer to 1, it represents a higher collaboration, and the farther it gets from 1, it represents a lower collaboration. As it is clear from figures 1 and 2, whether in EBSCO database or in Science Direct, the yearly CC values do not exceed 0.4. The maximum CC value belongs to year 2002 in both databases, with a value around 0.42, and the minimum value belongs to the year 2003 with a yearly CC value of around 0.25. These figures indicate that there appears to be a little collaboration among those who are involved in research and publishing on "digital libraries".

The cumulative CC value for the whole 5 year period was also calculated for each set of articles. The resulted figures, 0.3614 for the articles found in EBSCO database, and, 0.349 for articles found in Science Direct database, both, while very close again, are indicative of a fairly low collaboration.

3. Calculation of CC for the whole bulk of articles

Comparison of the two sets of articles as extracted from Ebsco and Science Direct databases, revealed that there were only 6 duplicate articles in both sets. The duplicates were excluded from the sets before they were combined to form a new set of 339 articles. Another round of calculations then was performed to work out CC values on this basis. Not necessary to mention that anonymous articles were pinpointed and excluded from the combination. Thus a total of 298 items remained in calculations, results of which are presented in Table 3.

As figures in Table 3 indicate, CC values show a growth of 38% in the year 2004 in comparison with the year 2000. However, all in all, neither each of the yearly CCs, nor the collective one are remarkable. This can be indicative of the fact that *collaboration* is perhaps not a matter of serious attention among those who are involved in writing/researching on "digital libraries" – although, as mentioned earlier, the degree of collaboration has grown, and is higher in 2004 than the previous years. Figure 3 illustrates the pattern of changes in CC values for digital libraries over the last 5 year period.

Table 3: Collective CC values For digital libraries in a 5 year period

Year	CC values
2000	0.2753
2001	0.305
2002	0.379
2003	0.3098
2004	0.3811
5 year period	**0.3283**

Figure 3 also presents the yearly status of collaboration within the mentioned 5 years. As it is clear from figure 3, the majority of articles are single-authored. Only a few articles are multi-authored. One can seldom find multi- and mega-authored[1] articles in this group. Also, as is shown in the figure, number of articles with 3 authors has raised in 2004 (series 5) in comparison with previous years. Cumulative CC values indicate the same trend.

Fig. 3: Pattern of changes in CC values over the last 5 years

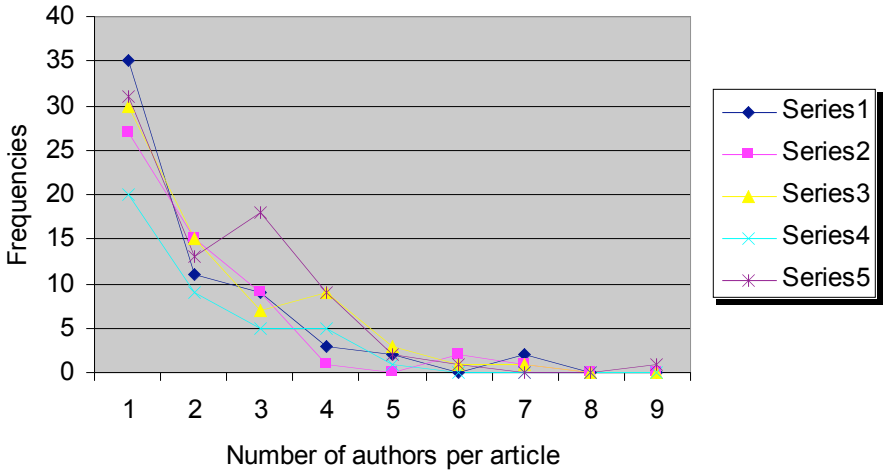
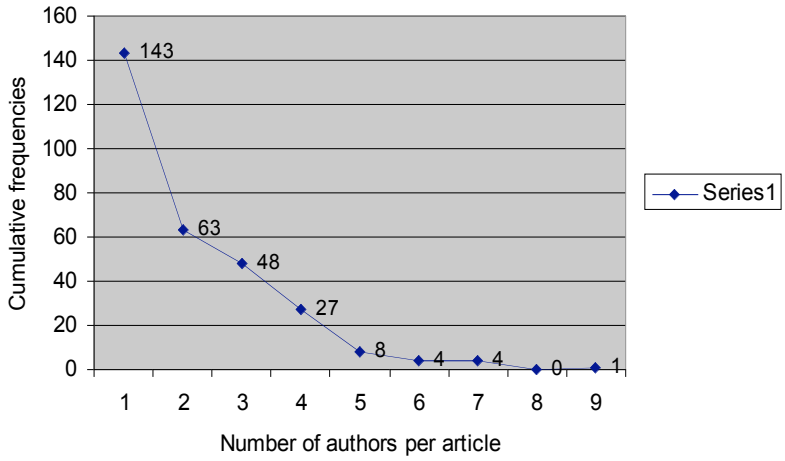


Figure 4 present the cumulative status of collaboration in the 5 years. As it is illustrated in Figure 4, the total number of single-authored articles form the majority, and there is a decline as the number of authors per article grows. The total number of articles with more than five authors fall at the bottom.

[1] . Refers to articles with more than 5 authors.

Fig.4: Cumulative representation of the number of authors per article



4. Collaboration in subject areas

In the total set of articles, 109 subjects were identified to be associated with digital libraries. The mean number of authors per article was calculated for each subject. The highest number of authors (7) is that of "digital library portals" and "Guidelines", and the lowest (1) belongs to 26 different among which are some applications of digital libraries and miscellaneous issues like "future of digital libraries", "conference reports", "metadata" and "resources". Overall, investigation reveals that the highest rate of collaboration lies in subjects that relate to *technicalities*, *applications, experiences, performance,* and issues associated with *information processing and retrieval*. Table 4 presents the identified subject areas and the related mean number of authors.

5. Conclusion

Nowadays, with development of modern multi-disciplinary subject areas, "digital libraries" included, one can rarely expect the single researcher/author to be capable of, and posses the expertise to deal with all aspects of a matter. Therefore, collaboration is one of the features of the present day academic and scientific activities. The more a community of scientists is collaborative, the more it is expected to be capable of reasonable growth and advancement. Examination of more than 300 articles published on digital libraries found in two full-text databases, indicate that authors/researchers in this area do not seem to be very much willing for collaborative authorship. This is well demonstrated in calculations of CC values both in each single set of articles, as well as in the combined set. If "digital libraries", as a discipline, is expected to enjoy a reasonable growth and advancement, its custodians, ie, authors and researchers in this field, are expected and required to develop a more collaborative morale, as well. The present study also revealed that, where collaboration exists in association with digital libraries, it is more in areas that deal with technicalities and experimentation in this field.

As the final note, it should be mentioned that the sample of 345 articles selected for examination in this study, are indicative only of the authorship aspect of a wide range of activities that have been undertaken in regard with digital libraries. There have been significant activities like Dublin Core, OAI, etc. that have certainly grew out of collaborative efforts of the scientists in practical realms, and should not be ignored.

Table 4. Identified subject areas and related Mean number of authors

Subject	Mean # of authors
DL, portals	7
Guidelines	7
Application in: Geography	6
DL, Children's	6
Networking	6
Application in: Communication	5
Character recognition	5
Experiments	4.5
Data management	4.3
Application in: Neuroscience	4
Application in: Physics	4
Application in: Weather and space	4
DL, Architecture	4
DL, Scholarly	4
Document management	4
Encoding techniques	4
Expert systems	4
Journal impact	4
Methods	4
Rural areas	4
Application in: Astronomy	3.75
Access	3.5
Fuzzy sets	3.5
Interoperability	3.5
Spatial issues	3.5
Application in: Medicine	3.25
Application, Gen.	3
Automation	3
Browsing	3
Cataloging	3
Classification	3
DL, Semantic	3
Euler Diagrams	3
Evaluation	3
Government information	3
Information processing	3
Information seeking	3
Information retrieval	3
Navigation	3
Psychology	3
Video compression	3
Video segmentation	3
Usage and user	2.87
Design	2.6

Search	2.6
DL, distributed	2.5
Performance issues	2.5
Publication & publishing	2.5
Representation	2.5
Application in: Education	2.33
Image retrieval	2.3
Indexing	2.3
Interface	2.25
Software	2.14
Application in: Geometry	2
Application in: Learning	2
Collection management	2
Compressed images	2
Computation methods	2
Computer graphics	2
Delphi study of	2
DL, Development	2
DL, Asian	2
Document retrieval	2
Information services	2
Infrastructure	2
Newspapers	2
Observation systems	2
Queries	2
Retrieval	2
Situations	2
Success and failure	2
Texture issues	2
Case studies	1.8
Cases	1.6
DL, use	1.6
Visualization	1.6
Experiences	1.4
Application in: Research	1.3
DL, construction	1.3
Projects	1.3
Collaboration	1.25
Patenting	1.25
Application in: Industry	1
Collection development	1
Commercialization	1
Conference report	1
Consortia	1
Content analysis	1
Copyright	1
Digitization	1
DL, models	1

DL, General Description	1
eResources	1
Future of	1
Image databases	1
Language understanding	1
Library use	1
Metadata	1
National Digital library	1
Opportunities	1
Organizational impact	1
Pattern recognition	1
Perspectives of	1
Presentation	1
Quality	1
Resources	1
Security	1
Subscription	1
Application in:	

References:

1. Ajiferuke, L; Burrel, Q.; Tague, J. (1988) Collaborative Coefficient: A single Measure of the Degree of Collaboration in Research. *Scientometrics*, 14: 421-433.
2. Havemann, F. (199?). Collaboration and Productivity of West-German Biomedical Researchers 1980-84 and 14 years later. Available at: http://141.20.126.8/~fhavem/Delhi.pdf.
3. Havemann, F. (?) Bibliometric indicators and their use for research evaluation: an analysis of highly productive biomedical teams. In: *Analysis of Highly Productive Biomedical Teams*.
4. Farajpahlou, A. and Samiromizadeh, M. A Comparative Study of Collaboration in Humanities and Engineering [unpublished article].