Information Processing and Management 51 (2015) 130–140

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Breadth and depth of citation distribution

Siluo Yang *, Ruizhen Han

School of Information Management, Wuhan University, Wuhan 430072, China

A R T I C L E I N F O

Article history:
Received 18 October 2013
Received in revised form 24 September 2014
Accepted 16 December 2014
Available online 7 January 2015

Keywords:
Citation analysis
Bibliometrics
Citation distribution
X index
Y index
XY index

A B S T R A C T

This study proposes a new 4D (i.e., spatial, temporal, breadth, and depth) framework for citation distribution analysis. The importance and differences in the breadth and depth of citation distribution are analyzed. Easily computable indices, *X*, *Y*, and *XY*, are proposed, which provide estimates of the breadth and depth of citation distribution. A knowledge unit can be an article, author, institution, journal, or a set of something. Index *X*, which represents the breadth of citation distribution, is the number of different knowledge units that cite special knowledge units. Index *Y*, which represents the depth of citation distribution, is the maximum number of citations among several knowledge units that refer to specific knowledge units. Index *XY*, which synthetically represents Indices *X* and *Y*, the feature and focus impacts of a knowledge unit, is index *X* divided by index *Y*. We analyze empirically the citation and reference distributions of 84 journals from the "Information science and library science" category of the Journal Citation Reports (2012) at the journal-to-journal level. Indices *X*, *Y*, and *XY* reflect the actual breadth and depth of citation distribution. Differences exist among Indices *X*, *Y*, and *XY*. Differences also exist between these indices and other bibliometric indicators. These indices cannot be replaced by existing bibliometric indicators. Specifically, the absolute values of indices *X* and *Y* are good supplements to existing bibliometric indicators. However, index *XY* and the relative values of Indices *X* and *Y* represent new aspects of bibliometric indicators.

1. Introduction

The first scientific journals appeared in the 16th century, but the systematic practice of citation was popularized later (Nicolaisen, 2007). With the rapid development of science and technology, researchers have begun to pay more attention to citations and references. In the 20th century, the practice of citing other works is almost second nature to anyone writing a scholarly or scientific paper (Kaplan, 1965). In the 1950s, Johns Hopkins University Librarian Eugene Garfield was inspired by an existing law database to create the Science Citation Index (SCI), which enables researchers to analyze millions of citations automatically and conveniently. Thereafter, some important contents and concepts of citation analysis were proposed, such as "Bibliographic Coupling" (Kessler, 1963), "Science Citation Network" (Price, 1965), "Co-Citation" (Small, 1973), and "Citation Visualization" (White & McCain, 1998). In 1998, the advent of the network version database Web of Science further promoted the popularization of citation analysis. In the network environment, Webometrics (Almind & Ingwersen, 1997) and Altmetrics (Wouters & Costas, 2012) were proposed, and a new era of the development of citation analysis began.

* Corresponding author. Tel.: +86 18040599879.
E-mail address: slyang2005@aliyun.com (S. Yang).

http://dx.doi.org/10.1016/j.ipm.2014.12.003

Currently, citation analysis is widely used in scientific evaluation, scholarly communication reveal, academic behavior analysis, and information retrieval (Garfield, 1983; Hammarfelt, 2011; Ketzler & Zimmermann, 2013).

Citation analysis is the method of examining the frequency, patterns, and graphs of citations in articles and books (Garfield, 1983). However, the development of citation analysis was accompanied by a controversy regarding its effectiveness and reliability. Citation analysis received criticisms that pertain to weak basic theory on citation, unknown citing motivation, deficient citing process, disadvantageous citation analysis methods and data, and limited citation application (Bornmann & Daniel, 2008; MacRoberts & MacRoberts, 2010; Nicolaisen, 2007). Frequency and distribution are key points in citation analysis. Most researchers focus on citation frequency analyses (Egghe, Bornmann, & Guns, 2011), such as the simple statistical analysis of citation frequency, citation weighted analysis involving Eigenfactors, and a combination of other factors, such as H-index combining citation frequency and paper number (Bergstrom, 2007; Price, 1965; Yang, Han, Ding, & Song, 2012). Specialized research on citation distribution consists of four main aspects. First, it consists of the analysis of the mechanism and curves of the citation distribution model in general (Redner, 1998; Rodríguez-Navarro, 2011; Sangwal, 2013). Second, it consists of spatial and temporal distribution (Larivière, Archambault, & Gingras, 2008), for example, impact factor is a result of combining spatial and temporal analyses. Ding, Liu, Guo, and Cronin (2013) studied specific location distribution of citations in citing literature context. Third, several scholars have studied information diffusion by analyzing citation networks or citation patterns (Chatterjee & Chowdhury, 2008; Shi, Tseng, & Adamic, 2009). The breadth of citations represent the diffusion of knowledge into other fields and from basic to applied research and development (R&D), which was initially suggested by W.F. Lancaster in his "issue management" studies (Lancaster & Lee, 1985). Wu (2013) studied the geographical knowledge diffusion and spatial diversity by exploring and investigating the spatial properties of citation distances and patterns. Fourth, some researchers have explored the breadth of citation distribution at the macro level. Evans (2008) argued that researchers cite recent papers and concentrate their citations on a few papers despite the availability of online research (older and recent). Larivière, Gingras, and Archambault (2009) challenged the conclusion of Evans by analyzing the changes in the concentration of citations received within 2 and 5 years by papers published between 1900 and 2005. Their results showed that the dispersion of citations is increasing. The conclusions of Yang, Ma, Song, and Qiu (2010) were consistent with those of Larivière et al. (2009) to a certain extent. However, few papers have studied systematically the breadth and depth of citation distribution at the micro level and lack indicators that measure the breadth and depth of citation. Thus, we should pay attention to these indicators to strengthen and validate citation distribution analysis (Yang et al., 2012).

2. 4D model of citation distribution

We argue that citation distribution can be divided into four dimensions, namely, spatial, temporal, breadth, and depth distribution. Citation analysis can also be studied in general regardless of the four dimensions, which integrate citations frequency statistics, citations weighted analysis, citations network analysis, and research citations related to the topic (see Fig. 1).

Spatial and temporal distribution are universal. Temporal distribution is the variation in distribution related to time. Everything has a life cycle, and year, month, and day are used as the units of analysis in bibliometrics. For example, the series of classic laws of literature is growing and aging. By contrast, spatial distribution is the variation in distribution related to position in space. In citation analysis, spatial distribution is not limited to geographical space and can be extended to the relationship between various knowledge units, such as country, organization, and journal. The distribution laws of

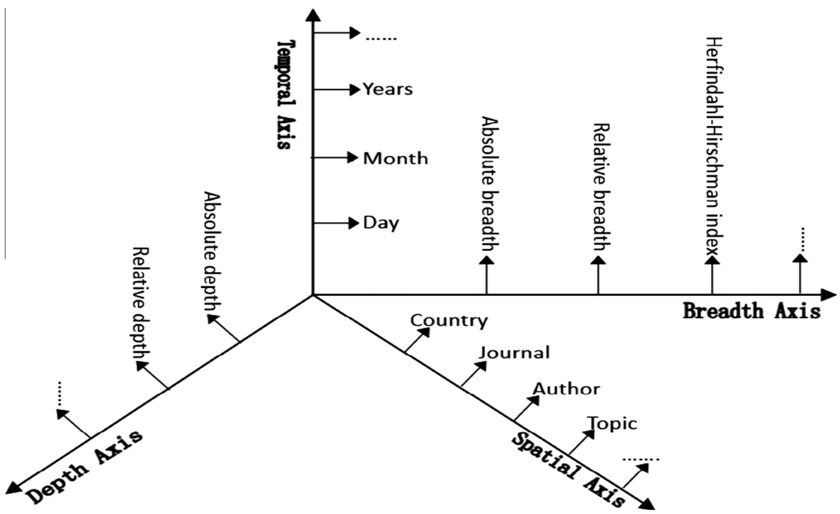

Fig. 1. 4D model of citation distribution.

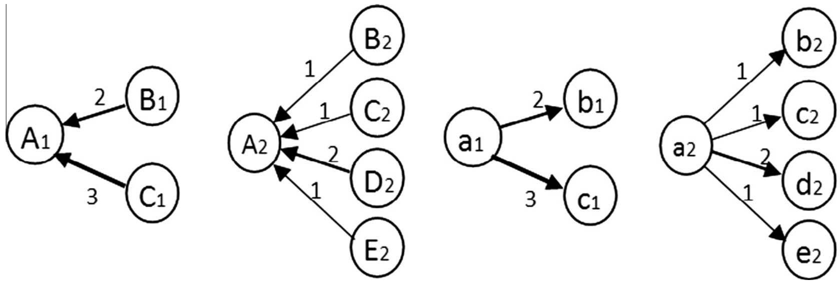

Fig. 2. Differences in the breadth and depth of citation distribution.

bibliometrics proposed by Bradford, Lotka, and Zipf belong to this type (Wilson, 2000). Furthermore, the networks of scientific papers can have a metaphor for topographic maps. According to Price (1963), "[with] such a topography established, one could perhaps indicate the overlap and relative importance of journals and, indeed, of countries, authors, or individual papers by the place they occupied within the map." In citation analysis, numerous indicators, such as impact factor, Eigenfactor, and H-index, can be used to measure spatial and temporal distribution.

Breadth and depth are significant factors of citation distribution. Several studies have investigated the curve represented by citation aggregation and dispersal (Zhang, 2013). As shown in Fig. 2, breadth and depth are different aspects of citation that can be represented through the width and number of lines or edges, respectively. The breadth and depth of citation distribution represent diffusion and concentration, respectively. Citation analysis uses citations in scholarly work to establish links to other works or researchers. This type of analysis is one of the most widely used bibliometric methods. A citation implies a relationship between parts of or the entire cited literature and a part of or the entire citing literature. The more an author or journal is cited in other scholarly articles, the stronger the author or journal's influence on a specific discipline. However, the effect of citing or being cited is affected by breadth and depth. For example, specific knowledge or scholars may influence deeply the knowledge structure of another scholar. The specific knowledge or scholars can be repeatedly cited by another scholar, thereby forming the depth of influence. However, with the development of inter-disciplines, scholars may widely adopt more extensive knowledge from various areas or academic knowledge, which form the width effect of citation. In the digital and network environment, an increasing number of journals and papers have been made available online by the digitization of historical archives and open access. All documents in the literature database have the same likelihood of being retrieved and accessed by scholars, and thus, numerous studies can be cited (Larivière et al., 2009). However, the articles referenced tended to be more recent as more journal issues came online, and as a result, fewer journals and articles were cited and more of those citations were to fewer journals and articles (Evans, 2008). Such a structure is responsible for the obvious differences in the breadth and depth of citation distribution.

3. Indices *X*, *Y*, and *XY*

The breadth and depth of citation distribution are easy to understand. However, the principal issue is the proper use of citation analysis as a tool in assessing such a situation and in describing the influence of breadth and depth. We believe citation analysis proposes breadth and depth as the relevant dimensions along which a citation distribution can be described quite conveniently by Indices *X*, *Y*, and *XY*.

3.1. Index X

Breadth is studied by analyzing the citation scope of a special knowledge unit in a certain scientific field, for example, an author or a journal. The concept of breadth hints that the knowledge unit influences the widening range of other knowledge. Index *X*, which represents the breadth of citation distribution, is the number of different knowledge units that cite special knowledge units. Index *X* is considered a special and independent measure of citation distribution. Index *X* concerns the number of different knowledge units, say, authors, that cite a specific author. The lines in Fig. 2 represent citation. Index *X* can be represented through the number of lines, whereas index *Y* can be represented through the maximum width of lines. To some extent, the breadth of citation distribution is similar to the existing measure of the in-degree/out-degree in citation network analysis. Citation network is directed, that is, edges point to one direction from one node to another. The nodes have two different degrees. One is the in-degree, which is the number of incoming edges, and the other is the out-degree, which is the number of outgoing edges. Accordingly, index *X* has two values, namely, index X_{in} and index X_{in}, which represent the right and left side in Fig. 2, respectively.

3.2. Index Y

Depth is studied by analyzing the degree to which a knowledge unit cites another knowledge unit. For example, an author cites another author, and a journal cites another journal. This concept hints at the degree of the effect of a specific knowledge unit on another knowledge unit. Index *Y*, which represents the depth of citation distribution, is the maximum number of citations among several knowledge units that refer to specific knowledge units. However, why is the maximum number of citation received/reference made by a specific knowledge unit (e.g., journal and author) a measure of depth? Why would *Y* be the maximum number of citations and not the sum? First, citation analyses have been based on how many times the papers were cited. If another paper cites the paper, the prestige and influence are transferred to the cited papers. The level of prestige and influence depends on the frequency of citation. Similarly, if an author frequently cites a special author, the latter deeply influenced the former. The maximum number of citation received/reference made by a specific unit, which represents the maximum effect of a knowledge unit. However, a special person is influenced by different knowledge in different degrees. More influence implies more citations. Second, scholars often pay more attention to the maximum number of citation than the sum in bibliometrics because the maximum value is more statistically significant. For instance, index H synthesizes the maximum number of citations and publications. The sum of citation/reference has been widely used, but it cannot distinguish and measure the breadth and depth of citation distribution. For example, Author B1 and Author C1 referenced Author A1 twice and thrice, respectively; Author B2, Author C2, Author D2, and Author E2 referenced Author A2 once, once, twice, and once, respectively (Fig. 2). Overall, the number of times authors A1 and A2 were cited is the same. However, the two authors differ in the breadth and depth of citation. Third, index *Y* represents well the citation distribution curve. We discuss the related content in the fourth paragraph in Section 4.2.1.

3.3. Index XY

Index *X* represents the influence scope of citation distribution, whereas index *Y* represents the influence degree or level. The influence of knowledge units has different characteristics, that is, some are good at breadth and some are good at depth. For example, a universal knowledge can be referenced through knowledge of more different kinds, whereas these citations of a specialized and particular knowledge are concentrated in a few knowledge units. How is statistics measured and conducted then? Index *XY*, which synthetically represents Indices *X* and *Y*, the feature and focus impacts of a knowledge unit, is index *X* divided by index *Y*. Indices *X* and *Y* represent different aspects that influence the knowledge units, which is based on the set of citing units and the number of citations that reference special knowledge units. If one has two numbers, one can divide them or multiply them or do anything else with them. We argue for dividing them because this indicates the ratio of breadth and depth. A stronger argument in addition to the argument is the shape of the cumulative citation distribution. *XY* combination (X/Y) makes good sense to citation analysis, and we can see this division indicates the feature of citation distribution. For example, Indices *X*, *Y*, and *XY* of A1 and A2 are 2, 3, 2/3 and 4, 2, 2, respectively (Figs. 2 and 3). The value of *XY* is the angle between the *Y*-axis and *XY*1 or *XY*2 in Fig. 3.

Several indices are available for measuring the breadth and depth of citation distribution, but Indices *X*, *Y*, and *XY* are suitable and easy. Similarly, we can study the breadth and depth from the perspective of reference or citing, such that *X* represents the breadth of reference distribution. We can also calculate the relative values of Indices *X*, *Y*, and *XY*. These values can be obtained by dividing the total number of knowledge units cited or citing. These indices include self-citations. If self-citations exist, they can also be normalized because they are well-known outliers.

- Indices *X*, *Y*, and *XY* are useful in measuring citation distribution. These indices can be applied to the productivity and effect of a group of scientists, such as a department, university, or country, as well as a scholarly journal or a paper. Indices *X*, *Y*, and *XY* are easy to obtain and represent a single number, which makes them easy to compare. These indices can be found in the Thomson ISI Web of Science database and can even be automatically calculated by citation index databases. For example, the *X* dimension (breadth) can be observed in the Web of Science database with the Citation Analytic Tool.
- The three proposed indices can be used to measure the effect of the work of an individual or a journal, which prevents certain disadvantages of previous indicators of citation analysis. For example, two individuals with a similar H-index

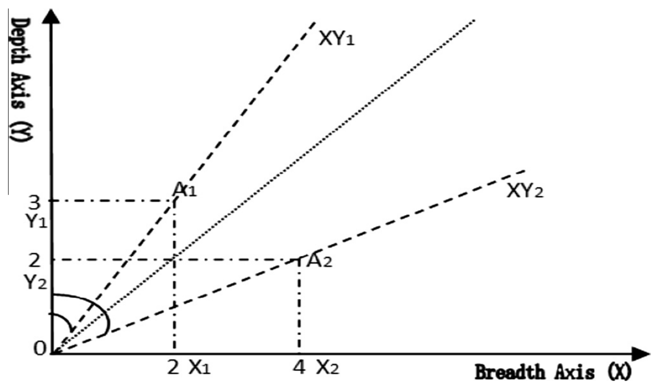

Fig. 3. Indices *X*, *Y*, and *XY* of citation distribution.

are comparable in terms of the calculating breadth and depth scientific influence, even if they have the same total number of citations. Conversely, in comparing two individuals with different total citation counts or H-index values, the individual with the higher H-index or total number of citations does not necessarily perform better in the breadth or depth influence. As Fig. 3 shows, knowledge units at different locations vividly demonstrate the depth and breadth of influence. Thus, certain professional knowledge units (journals or authors) have a relatively narrow effect, whereas others have wide effect. Certain knowledge units (authors) have far-reaching implications on specific areas. Others can have a significant effect on both breadth and depth, which can be easily calculated by index *XY*, as represented by the slope of A1 and A2 in Fig. 3.

- The breadth and depth of citation distribution have a wide range and can expand indices *X*, *Y*, and *XY* in numerous aspects. Knowledge units can comprise different analysis objects in bibliometrics. The width and breadth of citation can be analyzed between different knowledge units, such as scholar, organization, journal, and literature. They can also be analyzed between two papers, such as a paper cited several times by another. Breadth and depth can be combined with time, a measure of the absolute and relative relationship, and can be used to comparatively analyze similarities and differences from the perspective of cited and citing. We can also calculate the relative and absolute values of the breadth and depth of citation distribution. A comprehensive analysis that combines the four dimensions of spatial and temporal, as well as breadth and depth, can also be conducted with Indices *X*, *Y*, and *XY*.

4. Empirical analyses of breadth and depth of journal citation

In this section, we analyze empirically the breadth and depth of citation distribution through Indices *X*, *Y*, and *XY*. This objective can be achieved by addressing the following research questions:

Question 1: Can Indices *X*, *Y*, and *XY* reflect the actual breadth and depth of citation distribution?
Question 2: Are these breadth and depth indices of citation distribution different?
Question 3: Are these indices different from other bibliometric indicators? Can these indices be replaced by other bibliometric indicators?

4.1. Dataset

In principle, a knowledge unit can be an article, author, institution, or a set of something. However, we conduct our empirical study at the journal-to-journal level. Specifically, we analyze the citation distribution of journals citing other journals or being cited by different journals. A data set is obtained from Journal Citation Reports (JCR) 2012. We analyze the citation and reference distribution of 84 journals from the "Information science and library science" category. The reference distribution of 84 journals uses the Indices X_{in}, Y_{in}, and XY_{in}, whereas the citation distribution of 84 journals uses the Indices X_{out}, Y_{out}, and XY_{out}. In represents the absorption and input of knowledge, whereas *out* represents the influence and output of knowledge. These relative values of Indices *X*, *Y*, and *XY* can be obtained by dividing the total number of knowledge units cited or citing in Table 1. For example, Total Reference (4822) is the number of times that articles published in journals were cited in MIS QUART in 2012. The absolute value of X_{in} (1677) is the number of journals cited in MIS QUART in 2012. The relative values of X_{in} (0.35) is the absolute value of X_{in} divided by the Total Reference (1677/4822). The absolute value of Y_{in} (432) is the maximum number of citations among all journals cited in MIS QUART in 2012. The number of self-citation in MIS QUART in 2012 is also 432. The relative values of Y_{in} (0.09) is the absolute value of Y_{in} divided by the Total Reference (432/4822). Accordingly, Total Citation (7277) is the number of times that articles published in 2012 cited articles published in MIS QUART. XY_{in} (3.88) is X_{in}/Y_{in}; the relative values of XY_{in} equals the absolute value of XY_{in}. A large XY_{in} value indicates relative low concentration in the reference distribution of a journal. The maximum value of XY_{in} is 110.58 (LIBR J) in Table 1, which shows LIBR J referenced to a relatively large number of different journals but failed to focus fully on one or several journals.

We often analyze the correlation of two variables by measuring Spearman and Pearson correlation coefficients. The Pearson correlation coefficient is a measure of the strength of the linear relationship between two variables (Nikolić, Muresan, Feng, & Singer, 2012). Most of the relationships between variables in Table 1 are non-linear and follow a bivariate non-normal distribution. Hence, the Pearson correlation coefficient is unable to adequately represent the strength of the relationship between these variables. The Spearman correlation coefficient is a nonparametric measure of statistical dependence between two variables. It is appropriate for both continuous and discrete variables, including ordinal variables. Hence, we select Spearman correlation coefficient because we are mainly concerned about the numerical rank.

4.2. Results and discussion

4.2.1. Question 1

We analyze the cited and citing distribution of journals to validate whether Indices *X*, *Y*, and *XY* reflect the breadth and depth of citation distribution. The distribution of citations to scientific journals is extremely skewed (Price, 1965; Seglen, 1992). Based on JCR, a distribution curve that shows the sum of cumulative citations is a function of the amount of cited journals (Garfield, 1983), which illustrates that the centralized tendency of citation is more obvious than the paper

Table 1
Breadth and depth indices of journal citation in information and library sciences.

Rank	Abbreviated journal title	Impact factor	Citing (reference)						Cited					
			Total reference	Index X_{in}		Index Y_{in}		Index XY_{in}	Total citation	Index X_{out}		Index Y_{out}		Index XY_{out}
				Ab.	Re.	Ab.	Re.			Ab.	Re.	Ab.	Re.	
1	MIS QUART	4.659	4822	1677	0.35	432	0.09	3.88	7277	637	0.09	432	0.06	1.47
2	J INFORMETR	4.153	2221	740	0.33	362	0.16	2.04	943	165	0.17	222	0.24	0.74
3	J AM MED INFORM ASSN	3.571	6523	2184	0.33	798	0.12	2.74	5012	790	0.16	798	0.16	0.99
4	J INF TECHNOL	3.532	1696	862	0.51	114	0.07	7.56	1165	401	0.34	68	0.06	5.90
5	INFORM TECHNOL MANAG	3.025	2335	1013	0.43	230	0.10	4.40	247	79	0.32	134	0.54	0.59
6	ANNU REV INFORM SCI	2.174	*N*	*N*	*N*	*N*	*N*	*N*	380	150	0.39	70	0.18	2.14
7	SCIENTOMETRICS	2.133	8303	2914	0.35	1430	0.17	2.04	4555	528	0.12	1430	0.31	0.37
8	J HEALTH COMMUN	2.079	4549	1977	0.43	173	0.04	11.43	1614	504	0.31	173	0.11	2.91
9	INFORM SYST RES	2.01	4568	1618	0.35	278	0.06	5.82	3443	466	0.14	213	0.06	2.19
10	J AM SOC INF SCI TEC	2.005	9323	4166	0.45	813	0.09	5.12	4613	778	0.17	813	0.18	0.96
11	GOV INFORM Q	1.91	3902	2040	0.52	331	0.08	6.16	835	145	0.17	331	0.40	0.44
12	INT J INFORM MANAGE	1.843	2999	1367	0.46	141	0.05	9.70	970	340	0.35	94	0.10	3.62
13	J COMPUT-MEDIAT COMM	1.778	1445	807	0.56	86	0.06	9.38	1868	525	0.28	171	0.09	3.07
14	INT J COMP-SUPP COLL	1.717	1351	723	0.54	132	0.10	5.48	377	77	0.20	132	0.35	0.58
15	INFORM MANAGE-AMSTER	1.663	920	331	0.36	114	0.12	2.90	3091	484	0.16	116	0.04	4.17
16	INT J GEOGR INF SCI	1.613	5230	2402	0.46	338	0.06	7.11	2284	502	0.22	338	0.15	1.49
17	TELECOMMUN POLICY	1.594	3445	1868	0.54	428	0.12	4.36	961	208	0.22	428	0.45	0.49
18	EUR J INFORM SYST	1.558	3037	1234	0.41	267	0.09	4.62	1268	273	0.22	160	0.13	1.71
19	J STRATEGIC INF SYST	1.5	1909	646	0.34	163	0.09	3.96	743	222	0.30	133	0.18	1.67
20	J KNOWL MANAG	1.474	3576	1209	0.34	299	0.08	4.04	1392	282	0.20	299	0.21	0.94
21	LIBR INFORM SCI RES	1.4	1889	1109	0.59	100	0.05	11.09	536	136	0.25	71	0.13	1.92
22	INFORM ORGAN-UK	1.381	1203	579	0.48	59	0.05	9.81	261	104	0.40	37	0.14	2.81
23	INFORM SYST J	1.381	1243	592	0.48	78	0.06	7.59	696	223	0.32	75	0.11	2.97
24	SOC SCI COMPUT REV	1.303	1292	867	0.67	38	0.03	22.82	628	324	0.52	38	0.06	8.53
25	J MANAGE INFORM SYST	1.262	2156	788	0.37	152	0.07	5.18	2645	436	0.16	157	0.06	2.78
26	J INF SCI	1.238	1455	923	0.63	48	0.03	19.23	827	311	0.38	81	0.10	3.84
27	LEARN PUBL	1.182	293	206	0.70	27	0.09	7.63	208	68	0.33	27	0.13	2.52
28	MIS Q EXEC	1.143	164	116	0.71	15	0.09	7.73	189	71	0.38	19	0.10	3.74
29	J DOC	1.138	2123	1251	0.59	109	0.05	11.48	907	249	0.27	110	0.12	2.26
30	INFORM SOC	1.114	1201	923	0.77	23	0.02	40.13	561	271	0.48	24	0.04	11.29
31	RES EVALUAT	1.074	1157	666	0.58	80	0.07	8.33	411	112	0.27	67	0.16	1.67
32	KNOWL MAN RES PRACT	1.069	1451	813	0.56	60	0.04	13.55	240	89	0.37	60	0.25	1.48
33	J ASSOC INF SYST	1.048	2228	896	0.40	233	0.10	3.85	751	202	0.27	62	0.08	3.26
34	COLL RES LIBR	1.016	1018	549	0.54	67	0.07	8.19	441	95	0.22	67	0.15	1.42
35	J MED LIBR ASSOC	0.976	1169	655	0.56	95	0.08	6.89	627	248	0.40	95	0.15	2.61
36	LIBR RESOUR TECH SER	0.969	1020	491	0.48	53	0.05	9.26	131	32	0.24	53	0.40	0.60
37	ONLINE INFORM REV	0.939	2066	1098	0.53	92	0.04	11.93	480	176	0.37	92	0.19	1.91
38	J GLOB INF TECH MAN	0.917	620	339	0.55	35	0.06	9.69	117	58	0.50	27	0.23	2.15
39	J ACAD LIBR	0.885	1414	824	0.58	82	0.06	10.05	561	119	0.21	82	0.15	1.45
40	ETHICS INF TECHNOL	0.846	903	646	0.72	63	0.07	10.25	227	88	0.39	63	0.28	1.40
41	INFORM PROCESS MANAG	0.817	3185	1765	0.55	152	0.05	11.61	1681	446	0.27	158	0.09	2.82
42	INFORM TECHNOL PEOPL	0.767	1108	651	0.59	52	0.05	12.52	322	141	0.44	39	0.12	3.62
43	LIBR QUART	0.743	916	619	0.68	21	0.02	29.48	267	73	0.27	33	0.12	2.21
44	ELECTRON LIBR	0.667	1714	1009	0.59	87	0.05	11.60	257	91	0.35	87	0.34	1.05
45	HEALTH INFO LIBR J	0.662	996	561	0.56	88	0.09	6.38	364	168	0.46	88	0.24	1.91
46	LIBR HI TECH	0.621	1124	723	0.64	59	0.05	12.25	255	67	0.26	59	0.23	1.14
47	INFORM TECHNOL LIBR	0.595	*N*	*N*	*N*	*N*	*N*	*N*	92	43	0.47	11	0.12	3.91
48	SERIALS REV	0.524	528	360	0.68	23	0.04	15.65	109	51	0.47	23	0.21	2.22
49	INFORM RES	0.52	*N*	*N*	*N*	*N*	*N*	*N*	400	162	0.41	51	0.13	3.18
50	AUST ACAD RES LIBR	0.512	522	390	0.75	11	0.02	35.45	58	30	0.52	8	0.14	3.75
51	REV ESP DOC CIENT	0.453	792	522	0.66	37	0.05	14.11	88	24	0.27	32	0.36	0.75
52	J GLOB INF MANAG	0.452	1163	450	0.39	153	0.13	2.94	335	92	0.27	153	0.46	0.60
53	PROF INFORM	0.439	1705	1223	0.72	81	0.05	15.10	176	35	0.20	81	0.46	0.43
54	ASLIB PROC	0.432	1192	773	0.65	62	0.05	12.47	248	129	0.52	16	0.06	8.06
55	MALAYS J LIBR INF SC	0.423	615	374	0.61	69	0.11	5.42	66	35	0.53	13	0.20	2.69
56	KNOWL ORGAN	0.407	1128	846	0.75	48	0.04	17.63	102	33	0.32	48	0.47	0.69
57	PORTAL-LIBR ACAD	0.4	658	420	0.64	26	0.04	16.15	183	59	0.32	27	0.15	2.19
58	LIBR J	0.397	3702	3649	0.99	33	0.01	110.58	340	128	0.38	33	0.10	3.88
59	SCIENTIST	0.387	286	161	0.56	17	0.06	9.47	244	189	0.77	12	0.05	15.75
60	INFORM TECHNOL DEV	0.378	883	557	0.63	52	0.06	10.71	131	55	0.42	52	0.40	1.06
61	PROGRAM-ELECTRON LIB	0.377	884	634	0.72	19	0.02	33.37	179	98	0.55	17	0.09	5.76
62	INFORM DEV	0.375	761	567	0.75	18	0.02	31.50	74	37	0.50	18	0.24	2.06
63	RESTAURATOR	0.375	379	231	0.61	35	0.09	6.60	200	39	0.20	35	0.18	1.11
64	LIBRI	0.368	1172	839	0.72	19	0.02	44.16	125	64	0.51	13	0.10	4.92
65	DATA BASE ADV INF SY	0.341	564	363	0.64	32	0.06	11.34	323	143	0.44	12	0.04	11.92

(*continued on next page*)

Table 1 (*continued*)

Rank	Abbreviated journal title	Impact factor	Citing (reference)						Cited					
			Total reference	Index X_{in}		Index Y_{in}		Index XY_{in}	Total citation	Index X_{out}		Index Y_{out}		Index XY_{out}
				Ab.	Re.	Ab.	Re.			Ab.	Re.	Ab.	Re.	
66	ONLINE	0.341	31	30	0.97	2	0.06	15.00	138	100	0.72	10	0.07	10.00
67	J LIBR INF SCI	0.286	835	591	0.71	20	0.02	29.55	109	45	0.41	12	0.11	3.75
68	LIBR TRENDS	0.273	1521	1103	0.73	24	0.02	45.96	357	128	0.36	23	0.06	5.57
69	J SCHOLARLY PUBL	0.25	610	373	0.61	103	0.17	3.62	61	52	0.85	5	0.08	10.40
70	J ORGAN END USER COM	0.243	*N*	*N*	*N*	*N*	*N*	*N*	71	58	0.82	4	0.06	14.50
71	AUST LIBR J	0.239	657	530	0.81	11	0.02	48.18	45	21	0.47	11	0.24	1.91
72	INTERLEND DOC SUPPLY	0.231	408	269	0.66	37	0.09	7.27	65	53	0.82	12	0.18	4.42
73	LIBR COLLECT ACQUIS	0.185	316	195	0.62	20	0.06	9.75	100	24	0.24	26	0.26	0.92
74	CAN J INFORM LIB SCI	0.171	277	226	0.82	11	0.04	20.55	41	26	0.63	9	0.22	2.89
75	SOC SCI INFORM	0.167	1299	969	0.75	20	0.02	48.45	512	355	0.69	20	0.04	17.75
76	TRANSINFORMACAO	0.167	409	336	0.82	11	0.03	30.55	27	9	0.33	11	0.41	0.82
77	LIBR CULT REC	0.158	*N*	*N*	*N*	*N*	*N*	*N*	15	8	0.53	5	0.33	1.60
78	INFORM SOC-ESTUD	0.155	918	616	0.67	21	0.02	29.33	23	7	0.30	7	0.30	1.00
79	ECONTENT	0.127	*N*	*N*	*N*	*N*	*N*	*N*	28	20	0.71	5	0.18	4.00
80	AFR J LIBR ARCH INFO	0.111	362	310	0.86	6	0.02	51.67	28	20	0.71	5	0.18	4.00
81	PERSPECT CIENC INF	0.101	998	717	0.72	28	0.03	25.61	45	14	0.31	16	0.36	0.88
82	Z BIBL BIBL	0.07	234	187	0.80	11	0.05	17.00	9	5	0.56	5	0.56	1.00
83	INVESTIG BIBLIOTECOL	0.062	718	530	0.74	16	0.02	33.13	20	7	0.35	13	0.65	0.54
84	LIBR INFORM SC	0.05	41	40	0.98	2	0.05	20.00	4	3	0.75	2	0.50	1.50

Note. Ab. represents the absolute value; Re. represents the relative value. *N* represents the value missing in JCR.

distribution according to Bradford's Law. According to previous statistical data, Garfield conducted a deep exploration and proposed the famous conclusion that most references cited relatively few journals. In the 1969 SCI, only 25 journals (slightly more than 1% of the SCI coverage) were cited in 24% of all references, 152 journals were cited in 50% of all references, 767 journals were cited in 75% of all references, and approximately 2000 were cited in 85% of all references (Garfield, 1983).

We analyze these distributions of journals cited by the 84 journals from the "Information science and library science" category. We also analyze the distributions of these journals that cite the 84 journals. We find that the citation distribution is regular, uneven, and similar to the general description curve provided by Garfield (1983). On the one hand, a small number of journals obtain the vast majority of citations, which indicate a high degree of aggregation. On the other hand, citations are highly dispersed among numerous journals.

Eighty-four journals from the category "Information science and library science" are arranged in descending order by impact factor in 2012. *Information Processing & Management* is an information science journal and ranks 41st, whereas *Library Quarterly* is a library science journal and ranks 43rd. As illustrated in Figs. 4 and 5, we draw the accumulated citation distribution similar to Garfield's citation plot (Garfield, 1983). For example, the left graph in Fig. 4 represents the distribution of journals that cited the *Information Processing & Management* in 2012. The distribution of journal citation is clear. Approximately 9.4% of citations (158) are concentrated in a single journal (0.2%). Almost 30% citations are gathered from 6 journals (1.35%). That is, approximately 70% citations are provided by 19% of the journals. However, 30% citations (526) are spread in almost 360 journals (81%). The right graph in Fig. 4 represents the distribution of journals cited by *Information Processing & Management* in 2012. Statistical results show that the distribution of citations is uneven. The distribution curves of another 82 journals are similar to those presented in Figs. 4 and 5. These findings indicate that the citing and cited distribution of journal is skewed and regular to a certain extent.

As shown in Figs. 4 and 5, Indices *X*, *Y*, and *XY* can be used to clearly describe and explain the citation distribution. Index *X* is the number of different knowledge units that cite special knowledge units, as represented by the last point of *X* axis in Figs. 4 and 5. Index *Y* is the maximum number of citations among several knowledge units that refer to specific knowledge units, as represented by the first point of *Y* axis in Figs. 4 and 5. The first point of *X* axis and the last point of *Y* axis are 0 and 100, respectively; they are invariant. However, index *XY*, which is defined as index *X*/index *Y*, mainly reflects the shape of the curve. The combination of *X*, *Y*, and *XY* can further be used to analyze the breadth and depth distribution of citations. Fig. 6 reveals the citation distribution characteristic when considering Y_{in} and X_{in}, Y_{out}, and X_{out}, and XY_{in} and XY_{out}. X_{in} represents reference citing, whereas X_{out} represents cited. For example, the first coordinate in Fig. 6 has four quadrants. If the value of a journal is part of Quadrant A, the reference of the journal has high breadth and depth distribution. Furthermore, Indices *X*, *Y*, and *XY* can be diachronically analyzed to present the pattern and evolution of citation distribution in some journals.

Overall, Indices *X*, *Y*, and *XY* are easy to calculate. They reflect the actual breadth and depth of citation distribution.

4.2.2. Question 2

Table 1 and Fig. 7 show the differences among various indicators. Obvious differences exist between the indicators of breadth and width. Symmetric matrices are formed through the Spearman correlation analyses of each index. However, Fig. 7 does not show symmetric matrices. Correlation values exist between the absolute value of Indices *X*, *Y*, and *XY* and others indices in the upper-right part, whereas correlation values exist between relative values in the bottom-left part.

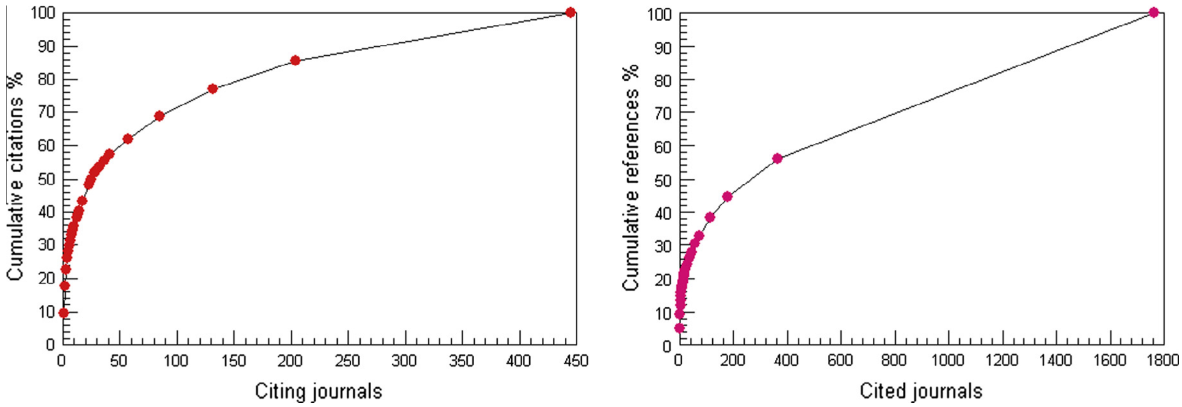

Fig. 4. Citation and reference distribution curves of *Information Processing & Management* in 2012.

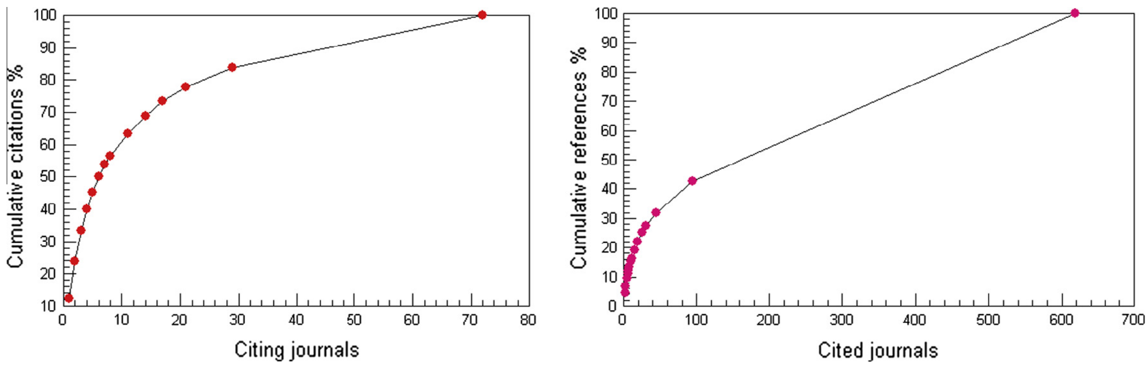

Fig. 5. Citation and reference distribution curves of *Library Quarterly* in 2012.

Fig. 7 presents the Spearman correlation among the indices of citation distribution. Spearman correlation represents the relation of the rank order. First, the absolute and relative values of indices XY_{in} and XY_{out} are not highly correlated with those of other indices. Most values exhibit a negative correlation. A low correlation (0.298) exists between XY_{in} and XY_{out}, although the relationship remains extremely strong, that is, the correlation is significant at the 0.01 level. Second, the absolute values of X_{in} is correlated with Y_{in} (0.689), whereas the relative values of X_{in} have negative correlation with Y_{in} (−0.678). The relative values of X_{in} are less closely correlated with indices XY_{in} (−0.169) and XY_{out} (−0.167). The absolute values of XY_{out} are low correlated with X_{in} (−0.26) and Yin (−0.281). Third, the correlation between the absolute values of X_{out} and Y_{out} is strong (0.762), whereas that between the relative values of X_{out} and Y_{out} is weak (−0.144). The absolute values of X_{out} is not correlated with indices XY_{out} (0.193). Fourth, relatively high correlation is also noted among the absolute values of Indices X_{in}, Y_{in}, XY_{in} and X_{out}, Y_{out}. The absolute values of Y_{out} is highly correlated with Y_{in} (0.908) mainly because, in the depth aspect, the maximum number a journal cited another and has been cited by another is often similar, particularly when the maximum number is derived from the same journal (i.e., journal self-citation). However, in the relative value, the correlation is low, and the relationship among Indices X_{in}, Y_{in}, XY_{in} and X_{out}, Y_{out}, XY_{out} is weak.

4.2.3. Question 3

Fig. 7 show high correlations among Impact Factor, Eigenfactor, Total reference, and Total citation, especially between Eigenfactor and total citation (Spearman rho = 0.962). This result is consistent with that of Davis (2008).

The Spearman correlation represents the relationship of the rank order in Fig. 7. First, the absolute values of Indices *X* and *Y* are correlated with those of other bibliometric indicators, such as impact factor, Eigenfactor, total reference, and total citation. All correlations are significant at the 0.01 level (two-tailed). Strong correlation is found among several indicators (Pearson rho > 0.9). For example, Eigenfactor are highly correlated with X_{out} and Total citation mainly because Total citation and X_{out} represent the overall effect and influence breadth of citation distribution, respectively. Eigenfactor indicates that prestige is measured by a weighting mechanism in the citation network, which maybe explains that the influence of breadth is correlated with prestige. In addition, total reference is highly correlated with X_{in} (0.95), which may be attributed to the fact that a journal with more references tends to cite various journals. However, most of the existing bibliometric indicators have low or negative correlation with indices *XY*.

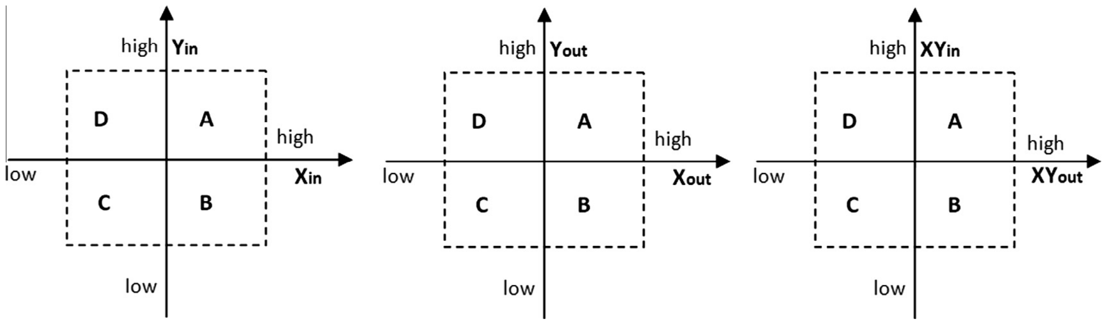

Fig. 6. Citation distribution analysis combining Indices *X*, *Y*, and *XY*.

Spearman's rho		Impact Factor	Eigenfactor	Total Reference	Index Xin	Index Yin	Index XYin	Total Citation	Index Xout	Index Yout	Index XYout
Impact Factor	Coefficient	1.000	.865**	.730**	.593**	.805**	-.657**	.841**	.765**	.852**	-.180
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.000	.100
	N	84	84	78	78	78	78	84	84	84	84
Eigenfactor	Coefficient	.865**	1.000	.796**	.699**	.781**	-.548**	.962**	.920**	.843**	-.004
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000	.000	.971
	N	84	84	78	78	78	78	84	84	84	84
Total Reference	Coefficient	.730**	.796**	1.000	.950**	.829**	-.388**	.824**	.748**	.843**	-.220
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.000	.053
	N	78	78	78	78	78	78	78	78	78	78
Index Xin	Coefficient	-.793**	-.721**	-.659**	1.000	.689**	-.169	.717**	.665**	.717**	-.167
	Sig. (2-tailed)	.000	.000	.000	.	.000	.140	.000	.000	.000	.145
	N	78	78	78	78	78	78	78	78	78	78
Index Yin	Coefficient	.519**	.414**	.213	-.678**	1.000	-.795**	.805**	.701**	.908**	-.333**
	Sig. (2-tailed)	.000	.000	.062	.000	.	.000	.000	.000	.000	.003
	N	78	78	78	78	78	78	78	78	78	78
Index XYin	Coefficient	-.657**	-.548**	-.388**	.848**	-.954**	1.000	-.565**	-.479**	-.664**	.298**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.000	.000	.008
	N	78	78	78	78	78	78	78	78	78	78
Total Citation	Coefficient	.841**	.962**	.824**	-.751**	.425**	-.565**	1.000	.959**	.878**	-.012
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.	.000	.000	.912
	N	84	84	78	78	78	78	84	84	84	84
Index Xout	Coefficient	-.622**	-.621**	-.621**	.653**	-.430**	.553**	-.636**	1.000	.762**	.193
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.	.000	.079
	N	84	84	78	78	78	78	84	84	84	84
Index Yout	Coefficient	-.185	-.420**	-.136	.089	.074	-.021	-.418**	-.144	1.000	-.426**
	Sig. (2-tailed)	.091	.000	.236	.440	.522	.857	.000	.191	.	.000
	N	84	84	78	78	78	78	84	84	84	84
Index XYout	Coefficient	-.180	-.004	-.220	.260*	-.281*	.298**	-.012	.619**	-.841**	1.000
	Sig. (2-tailed)	.100	.971	.053	.022	.013	.008	.912	.000	.000	.
	N	84	84	78	78	78	78	84	84	84	84

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Fig. 7. Spearman correlations among the indices of citation distribution. Correlation values between the absolute values of Indices *X*, *Y*, and *XY* and other indices are shown in the upper-right part; correlation values between the relative values of Indices *X*, *Y*, and *XY* and other indices are shown in the bottom-left part.

Second, the relative values of Indices *X*, *Y*, and *XY* are weakly correlated with those of other bibliometric indicators, and some of these indices exhibit negative correlation. Some relationships are also weak because the correlations are insignificant at the 0.01 or 0.05 level.

5. Conclusion

This study reports that citation distribution can be generally divided into four dimensions, namely, spatial, temporal, breadth, and depth distribution. We introduce a 4D model of citation distribution. The significance and differences in the breadth and depth of citation distribution are analyzed. With Indices *X* and *Y*, index *XY* is proposed to measure the breadth and depth of citation distribution. Index *X*, which represents the breadth of citation distribution, is the number of different knowledge units that cite special knowledge units. Index *Y*, which represents the depth of citation distribution, is the maximum number of citations among several knowledge units that refer to specific knowledge units. Index *XY*, which synthetically represents indices *X* and *Y*, the feature and focus impacts of a knowledge unit, is calculated as index *X* divided by index

Y. Indices X, Y, and XY can be used for the statistical analysis of absolute and relative values in citation and reference distribution.

We analyze empirically the citation and reference distribution of 84 journals under the "Information science and library science" category. Indices X, Y, and XY reflect the actual breadth and depth of citation distribution. Differences exist among indices X, Y, and XY, as well as between these indices and other bibliometric indicators. Thus, these indices cannot be replaced by other bibliometric indicators. Specifically, the absolute values of indices X and Y are a good supplement to other bibliometric indicators. They are correlated with most of the other bibliometric indicators, such as impact factor, Eigenfactor, total reference, and total citation. However, index XY and the relative values of indices X and Y are a new aspect of other bibliometric indicators. They have low and even negative correlations with other bibliometric indicators. Certain relationships between these indices are insignificant at the 0.01 and 0.05 level.

In summary, this study proposes the use of easily computable Indices X, Y, and XY to estimate the breadth and depth of citation distribution. However, further studies must be conducted to verify our results. In the next stage, we intend to conduct the following:

(1) In-depth study at the theoretical level: This study will include citation distribution mechanism, citation motive, validity, advantages and disadvantages of these indices, and the citation distribution relationship between breadth and depth. Citation distributions involves many dimensions. The empirical analysis and theoretical models have to be studied significantly.
(2) Empirical study at different levels: This study will consider the numerous knowledge units in bibliometrics such as author, institution, journal, and paper. The breadth and depth of citation distribution can be analyzed based on these aspects combining synchronically and diachronically analysis, for example, the citation distribution of documents citing or cited by different authors in different time period. We will also analyze Indices X, Y, and XY in network environment, which are expanded to Webmetrics or Altmetrics.
(3) Optimization or improvement of these indices: H-index has gained widespread recognition. To solve one or more disadvantages of H-index, different researchers have proposed numerous variants, such as g-index, c-index, s-index, and m-index. Similarly, Indices X, Y, and XY must be modified to emphasize various features. For example, the index min(X, Y) synthetically may represent the significance of a knowledge unit defined as Min(X, Y). In addition, Index X_{in}/X_{out} may represent the relationship between absorption and output of knowledge units.

Acknowledgments

This research is funded by the Humanities and Social Science Foundation of the Ministry of Education of China (13YJCZH047) and A Foundation for the Author of National Excellent Doctoral Dissertation of PR China (2014094). We wish to thank two anonymous referees for important insightful comments and suggestions.

References

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'. *Journal of Documentation, 53*(4), 404–426.
Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *College & Research Libraries News, 68*(5), 314–316.
Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45–80.
Chatterjee, K. & Chowdhury, A. (2008). *Citations, networks and diffusions of knowledge: An economic analysis*. <http://econ.la.psu.edu/papers/Citations_avantika_08.pdf>.
Davis, P. M. (2008). Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *Journal of the American Society for Information Science and Technology, 59*(13), 2186–2188.
Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics, 7*, 583–592.
Egghe, L., Bornmann, L., & Guns, R. (2011). A proposal for a first-citation-speed-index. *Journal of Informetrics, 5*(1), 181–186.
Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science, 321*(5887), 395–399.
Garfield, E. (1983). *Citation indexing – Its theory and application in science, technology and humanities*. Philadelphia: ISI Press (pp. 12–56).
Hammarfelt, B. (2011). Citation analysis on the micro level: The example of Walter Benjamin's Illuminations. *Journal of the American Society for Information Science and Technology, 62*(5), 819–830.
Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation, 16*(3), 179–184.
Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10–25.
Ketzler, R., & Zimmermann, K. F. (2013). A citation-analysis of economic research institutes. *Scientometrics, 95*(3), 1095–1112.
Lancaster, F. W., & Lee, J. (1985). Bibliometric techniques applied to issues management: A case study. *Journal of the American Society for Information Science, 36*(6), 389–397.
Larivière, V., Archambault, E., & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology, 59*(2), 288–296.
Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology, 60*(4), 858–862.
MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology, 61*(1), 1–13.
Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 13*, 609–641.
Nikolić, D., Muresan, R. C., Feng, W., & Singer, W. (2012). Scaled correlation analysis: A better way to compute a cross-correlogram. *European Journal of Neuroscience, 35*(5), 742–762.
Price, D. J. D. (1965). Networks of scientific paper. *Science, 149*(3683), 510–515.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B, 4*, 131–134.
Rodríguez-Navarro, A. (2011). A simple index for the high-citation tail of citation distribution to quantify research performance in countries and institutions. *PLoS ONE, 6*(5), e20510.
Sangwal, K. (2013). Comparison of different mathematical functions for the analysis of citation distribution of papers of individual authors. *Journal of Informetrics, 7*(1), 36–49.
Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43*(9), 628–638.
Shi, X., Tseng, B., & Adamic, L. (2009). *Information diffusion in computer science citation networks*. 2009arXiv0905.2636S.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.
White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.
Wilson, C. S. (2000). Informetrics. *Annual Review of information Science and Technology, 34*, 107–286.
Wouters, P. & Costas, R. (2012). Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation. <http://www.cwts.nl/pdf/cwts-wp-2014-001.pdf> Retrieved 16.09.14.
Wu, J. (2013). Geographical knowledge diffusion and spatial diversity citation rank. *Scientometrics, 94*(1), 181–201.
Yang, S., Han, R., Ding, J., & Song, Y. (2012). The distribution of Web citations. *Information Processing & Management, 48*(4), 779–790.
Yang, S., Ma, F., Song, Y., & Qiu, J. (2010). A longitudinal analysis of citation distribution breadth for Chinese scholars. *Scientometrics, 85*(3), 755–765.
Zhang, C. (2013). The h-index, effectively improving the h-index based on the citation distribution. *PLoS ONE, 8*(4), e59912.