Authorship Pattern of Big Data Research Output: A Scientometric Analysis

R. Balasubramani
Assistant Professor
Dept. of Library and Information Science
Bharathidasan University
Trichy, Tamil Nadu.
e-mail:lisbala@gmail.com

G. T. Kohila
Research Scholar
Dept. of Library and Information Science
Bharathidasan University
e-mail: koela15@gmail.com

Abstract

The main objective of this study is to analyze the Authorship pattern in big dataresearch output.It aims to analyse the performance of scientists in contributing research output either individually or collectively. The bibliometric analytical technique was used to examine this topic for a period of 1998-2017.This study has observed a total of 10641 publications in big data research over a period of twenty years. Overall, 27900 authors contributed in 3193 journals with 10641 records of the publications from 7463 number of institutions that were located in 109 numbers of countries.

Keywords

Bigdata; Publications; Research; Bibliometrics; Analysis

Electronic access

The journal is available at www.jalis.in

Journal of Advances in Library and Information Science
ISSN: 2277-2219 Vol. 7. No.1. 2018. pp.15-19

1. INTRODUCTION

Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyse it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes.In the digital era, large amounts of data have become available by single click to researchers, policy makers and general public. Big data refers to datasets that are not only big, but also large in variety and velocity, which makes them difficult to handle using traditional methods. Due to the rapid growth of such data, quick fix need to be studied and provided in order to handle and extract value and knowledge from these datasets.It indicates the growing popularity of collaborative research endeavour among the scientists. Generally nowadays, research is carried out by group of researchers rather than by a single individual.

2. REVIEW OF LITERATURE

The scientific output of a country is measured by evaluating institutions or individual scientists. Two important parameters are examined, including over-all production and impact of scientific publications (Bornmann 2011; Cronin 1984; Franceschini et al.2007). The following three approaches are applied to evaluate these parameters using bibliometric indicators: i) counting the publication number; ii) counting the citation number; or iii) combining the first two counts to create hybrid indicators. Publication and citation counts are traditionally employed to indicate the influence or impact an author has within the research community (Adams 1990; Abramo&D'Angelo 2011, Wildegaard 2015).
The term "Big Data" has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time(Kubick, W.R. 2012). Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics based on large data samples reveals and leverages business change. However, the larger the set of data, the more difficult it becomes to manage (Russom, P. 2011).

3. OBJECTIVE

The main objective of this study is to analyse the growth of the research in bigdata that has arisen over the last few years. The study aims to evaluate the authorship pattern globally.

I. To find out the Authorship Pattern of article productivity
II. To Analyze the Year wise Distribution of Authorship Pattern of article productivity
III. To find out the Degree of collaboration of big data research output
IV. To analyze the Ranking of authors based on their research papers published.

4. DESIGN/METHODOLOGY/APPROACH

We have analysed publications of journals between published between 1998 to 2017. Furthermore, we have analysed the relationship with the authorship pattern and applied Lotka's Law of Author Productivity.

5. ANALYSIS AND INTERPRETATION

This study has observed a total of 10641 publications in big data research over a period of twenty years from 1998 to 2017 and indexed by Web of Science online Database indicates that the publication output in India. Overall, 27900 authors contributed in 3193 journals with 10641 records of the publications from 7463 number of institutions that were located in 109 numbers of countries.

Table 1: Authorship Pattern of article productivity

Authorship pattern	No. of Contri-bution	Cumulative of contribution	%	Cumula tive %
1	2089	2089	19.63	19.63
2	2114	4203	19.87	39.50
3	2018	6221	18.96	58.46
4	1616	7837	15.19	73.65
5	1095	8932	10.29	83.94
6	641	9573	6.02	89.96
7+	1068	10641	10.04	100.00
Total	10641		100.00	

The highest productivity of publications output (19.87%) from two authors. This is followed by single authors' contribution (19.63%).It is followed by three author contribution(18.96%).The least number of contribution is produced by six authors(6.02%).

5.1 AUTHORSHIP PATTERN

Table 2: Year wise Distribution of Authorship Pattern of article productivity during 1998-2017

No. of articles/Year		**Single**	**Double**	**Triple**	**Four**	**Five**	**Six**	**Seven & More**	**Total**
1998	articles	0	0	0	0	0	0	1	1
1999	articles	0	0	0	0	0	1	0	1
2000	articles	0	0	0	0	0	0	0	0
2001	articles	0	0	0	0	0	0	0	0
2002	articles	0	0	0	1	0	0	0	1
2003	articles	0	0	0	1	0	1	0	2
2004	articles	2	0	0	0	0	0	0	2
2005	articles	0	0	0	0	0	0	0	0
2006	articles	1	1	0	0	0	1	0	3
2007	articles	1	0	0	0	0	0	0	1
2008	articles	8	1	0	0	0	0	2	11
2009	articles	2	2	0	1	1	0	0	6
2010	articles	1	1	1	0	1	0	1	5
2011	articles	4	3	5	5	2	2	1	22
2012	articles	44	19	9	3	8	3	6	92
2013	articles	122	82	63	44	27	20	37	395

2014	articles	239	203	200	146	75	38	107	1008
2015	articles	465	441	389	280	205	130	201	2111
2016	articles	590	636	636	526	338	183	332	3241
2017	articles	610	725	715	609	438	262	380	3739
Total		**2089 (19.63)**	**2114 (19.87)**	**2018 (18.96)**	**1616 (15.19)**	**1095 (10.29)**	**641 (6.02)**	**1068 (10.04)**	**10641**

Above Table explicates the authorship pattern of contributions. Out of 10641 articles, a single author has contributed 19.63 percent, 19.87 percent of articles were published by two authors, 18.96 percent of articles were published by three authors, 15.19 % of the contributions were published by four authors, 10.29 percent of the contributions were published by five authors, 6.02 per cent of the contributions were published by six authors, 10.04 percent of articles were published from seven and more than seven authors group respectively.

5.2 REVEALS SINGLE VS MULTI AUTHORED AND DEGREE OF COLLABORATION IN BIG DATA RESEARCH OUTPUT

Table 3: Single Vs Multi Authored and Degree Of Collaboration

Year	Single authors		Multiple authors		Total	Degree of collaboration
	No. of output	%	No. of out put	%		
1998	0	0.00	1	0.01	1	1.00
1999	0	0.00	1	0.01	1	1.00
2000	0	0.00	0	0.00	0	0
2001	0	0.00	0	0.00	0	0
2002	0	0.00	1	0.01	1	1.00
2003	0	0.00	2	0.02	2	1.00
2004	2	0.10	0	0.00	2	0.00
2005	0	0.00	0	0.00	0	0
2006	1	0.05	2	0.02	3	0.67
2007	1	0.05	0	0.00	1	0.00
2008	8	0.38	3	0.04	11	0.27
2009	2	0.10	4	0.05	6	0.67
2010	1	0.05	4	0.05	5	0.80
2011	4	0.19	18	0.21	22	0.82
2012	44	2.11	48	0.56	92	0.52
2013	122	5.84	273	3.19	395	0.69
2014	239	11.44	769	8.99	1008	0.76
2015	465	22.26	1646	19.25	2111	0.78
2016	590	28.24	2651	31.00	3241	0.82
2017	610	29.20	3129	36.59	3739	0.84
Total	2089 (19.63)	100.00	8552 (80.37)	100.00	**10641**	0.80

A study of data from the above table indicates the degree of collaboration in research output of big data. The degree of collaboration is 0.80 during the study period 1998to 2017.i.e. out of the total 10641 articles ,80.37 percentages of them are published under joint author publication in big data research output. Averagely the degree of collaboration is 0.80

5.3 ANALYSIS OF HIGHLY PROLIFIC AUTHORS IN BIG DATA RESEARCH OUTPUT

We have tried to find the highly prolific authors, based on research output data for big data during the period of 1998 to 2017. The authors are ranked based on their number of research papers published.

Showing Highly Prolific Authors Based on Their Number of Research Papers

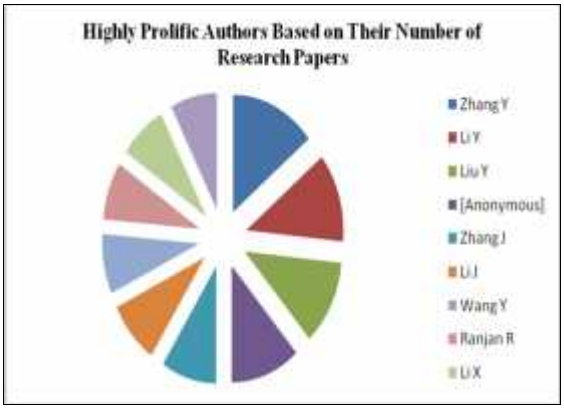
Table 4: Highly Prolific Authors In Big Data Research Output

S.no	Author	Recs	Percent	TLCS	TLCS/t	TLCSx	TGCS	TGCS/t	TLCR
1	Zhang Y	60	0.6	102	48.9	69	409	175.65	101
2	Li Y	52	0.5	42	23.17	27	244	105.55	62
3	Liu Y	50	0.5	69	31.43	51	252	103.58	73
4	[Anonymous]	49	0.5	6	1.5	6	3	0.75	0
5	Zhang J	39	0.4	31	10.75	22	181	68.42	38
6	Li J	37	0.3	22	10.33	19	140	61.83	62
7	Wang Y	35	0.3	15	6.25	14	66	27.75	31

8	Ranjan R	34	0.3	179	57.08	145	412	136	44
9	Li X	33	0.3	21	10	15	127	64.17	29
10	Liu J	33	0.3	20	6.15	13	85	27.98	42

The above table shows the details about the highly prolific 10 authors, the number of records, total local citation score, total global citation score and average citation per paper of big data research output during the period of 1998 to 2017. In this period27900 researchers have published 10641 research documents distributed over 3193 journals. The authors who have published 50 and above research documents are listed above, they are ranked on the basis of their number of papers published. The highest10 prolific authors together published a total number of 422 research documents.

Among the 27900 authors, "Zhang Y" hold the 1st rank and the author published 60 research papers with 102 total local citation scores 409 total global citation scores. Next to the author of "Li Y " hold the 2nd rank and the author published 52 research papers with 42 total local citation scores, 244 total global citation scores. The 3rd rank hold by "Liu Y" he contributed 50 research papers with 69 total local citation scores, 252 total global citation scores.



5.4 LOTKA'S LAW OFAUTHOR PRODUCTIVITY

Generally author productivity is determined on the basis of number of papers contributed by the big data in a specific field. It is quite relevant to study the impact of Lotka"s Lawin examining the author productivity in big data research. Below table presents the results o author productivity based on Lotka"sLaw.

Table 4: Lotka's Law of Author Productivity

No.of authors	Observed Number of authors with 'n'or (an)or(f)	Observed Percentage of authors 100xan/a1	Expected number of authors (an=an/n^2) or (p)	(F-P)^2/P
1	2089	100.00	2089.00	0.00
2	2114	101.20	528.50	4756.50
3	2018	96.60	224.22	14350.22
4	1616	77.36	101.00	22725.00
5	1095	52.42	43.80	25228.80
6	641	30.68	17.81	21811.81
7	319	15.27	6.51	14999.51
8	219	10.48	3.42	13581.42
9	125	5.98	1.54	9876.54
10	78	3.73	0.78	7644.78
11	57	2.73	0.47	6783.47
12	45	2.15	0.31	6390.31
13	39	1.87	0.23	6513.23
14	34	1.63	0.17	6596.17
15	18	0.86	0.08	4014.08
16	20	0.96	0.08	5080.08
17	13	0.62	0.04	3731.04
18	9	0.43	0.03	2898.03
19	9	0.43	0.02	3231.02
20	9	0.43	0.02	3582.02
21	4	0.19	0.01	1756.01
22	7	0.34	0.01	3374.01
23	4	0.19	0.01	2108.01
24	4	0.19	0.01	2296.01
25	2	0.10	0.00	1246.00
26	1	0.05	0.00	674.00
27	3	0.14	0.00	2181.00
28	2	0.10	0.00	1564.00
29	3	0.14	0.00	2517.00
30	3	0.14	0.00	2694.00
31	1	0.05	0.00	959.00
32	4	0.19	0.00	4088.00
34	5	0.24	0.00	5770.00

35	2	0.10	0.00	2446.00
36	2	0.10	0.00	2588.00
38	4	0.19	0.00	5768.00
41	1	0.05	0.00	1679.00
42	1	0.05	0.00	1762.00
44	4	0.19	0.00	7736.00
45	1	0.05	0.00	2023.00
48	4	0.19	0.00	9208.00
49	3	0.14	0.00	7197.00
50	1	0.05	0.00	2498.00
54	2	0.10	0.00	5828.00
57	2	0.10	0.00	6494.00
59	1	0.05	0.00	3479.00
73	1	0.05	0.00	5327.00
93	1	0.05	0.00	8647.00
101	1	0.05	2114.00	2112.00
	10641	509.38s	5132.13	289813.1

The above table shows that the testing of a prediction depends upon several factors. In the present analysis, the productivity is observed which is affected by several factors. If complete publication details of authors are taken, Lotka's law testing may present a different area. Moreover, if sampling size increases the results may get changed. From the above table it is evident that the expected values are not close to the observed values up to number of papers. Chi-square test are applied to verify the applicability of Lotka's law of scientific productivity. The statistical tests show that the Lotka's law in its generalized form does not fit the author productivity distribution pattern prepared for the straight count and for the contribution of complete study.

6. CONCLUSION

Big data attracts a lot of inter disciplinary attention and is a rapidly developing field of research. The present study is related to scientometrics analysis of research output of Big data from 1998-2017. The study has analysed authorship pattern, degree of collaboration among the authors. It is observed a total of 10641 publications in big data research over a period of twenty years from1998-2017and indexed by Web of Science online Database. Overall, 27900 authors contributed in 3193 journals with 10641 records of the publications from 7463 number of institutions that were located in 109 numbers of countries. Among the 27900 authors, "Zhang Y" hold the 1st rank and the author published 60 research papers with 102 total local citation scores 409 total global citation scores. The highest productivity of publications output (19.87%) from two authors. Out of 10641 articles, a single author has contributed 19.63 percent. The degree of collaboration is 0.80 during the study period 1998to 2017. The statistical tests show that the Lotka's law in its generalized form does not fit the author productivity distribution pattern prepared for the straight count and for the contribution of complete study.

REFERENCES

[1]. Abramo G., D'Angelo C, 2011. Evaluating research: from informed peer review to bibliometrics. *Scientometrics* 87 (3): 499-514.

[2]. Abramo G., D'Angelo C. A., 2011. National-scale research performance assessment at the individual level. *Scientometrics* 86(2): 347–364.

[3]. Amsaveni, N., & Ramesh, K. (2016). Mapping of Research Productivity in Forensic Science: A Scientometric Analysis. *Journal of Advances in Library and Information Science*, *5*(4), 372-378.

[4]. Jancy, S., Hariharan, R., & Subramanian, N. (2013). A Scientometric Study on Indian Journal of Pure and Applied Physics (2008-2012). *Journal of Advances in Library and Information Science*, *2*(3), 118-123.

[5]. Kubick, W.R. 2012. Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28.

[6]. Navalur, S. A., Balasubramani, R., & Kumar, P. A. (2012). Mapping of Global Research Performance in E-learning: A Scientometrics Analysis. *Journal of Advances in Library and Information Science*, *1*(3), 130-137.

[7]. Rakhi, V. S., & Nagarajan, M. (2013). Are the business schools in India doing their business? An Evaluation of IIM's Research productivity profile. *Journal of advances in library and Information science*, *2*(1), 25-30.

[8]. Russom, P. 2011. Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40

[9]. Surulinathi, M., Balasubramani, R., & Kalisdha, A. (2013). Continent wise Analysis of Green Computing Research: A Scientometric Study. *Journal of Advances in Library and Information Science*, *2*(1), 39-44.