Annals of Library and Information Studies
Vol. 63, June 2016, pp. 126-131

An evaluation of author productivity in artificial neural networks research in India during 1991-2014

Suresh Kumar

CSIR-National Institute of Science Technology and Development Studies, Dr. K S Krishnan Marg, (PUSA) New Delhi-110012, India, Email: sureshkr@nistads.res.in

Received: 02 December 2015; revised: 29 April 2016; accepted: 19 May 2016

The study examines the conformity of Lotka's law to authorship distribution in the field of Artificial Neural Networks research (ANNs) in India during 1991–2014 using Science Citation Index-Expanded. There were 3411 articles contributed by 5654 unique authors. Lotka's law was tested using methodology suggested by Pao and compared with maximum likelihood method advocated by Nicholls. The main elements involved in fitting in Lotka's law were identified. These includes criterion for taking a certain pair of observed data points for calculating Lotka's gradient, the constant for measurement of single author productivity and assessing goodness-of-fit. The results suggested that author productivity distribution, predicted by the modified Lotka's Law suggested by Pao, was confirmed to the ANNs discipline in India whereas methodology suggested by Nicholls was not able to explain the author productivity distribution for the same. Evaluation of the prolific authors indicated that most of them are among the top position in their respective institutions. However, they were not listed as first author in their publications supporting that all the authors should be considered while analysing author productivity.

Keywords: Lotka Law; Bibliometrics; Artificial Neural Networks, India.

Introduction

Growth of a specific field of research is reflected by efforts of both individuals and teams of professionals to share their findings through publication using all channels for communication such as journals, conferences etc. It ensures timely critical appraisal which represents the final stage in the research process. In this regard evaluation of author productivity has been used as proxy to indicate maturity of scholars in a particular field of research[1]. Lotka[2] (1926) was first to study author productivity from empirical data and proposed an inverse-square law relating to authors productivity distribution to understand worth of authors in progress of a subject field. He generated a frequency table of the number of authors and their contributions in the decennial index of Chemical Abstracts, 1907-1916, taking only authors with initial A and B. Similar, data were also collected from the name index of Auerbach's Geschichtstafeln der Physik. Lotka plotted, on a logarithmic scale, the number of authors against the number of contributions made by each author and found that in each case the points were closely scattered about a straight line having a slope of approximately two. On the basis of these data sets, deduced that the number of authors making x contribution is about $1/x^n$ of those making one contribution, n is often close to a value of 2 implying that about 60% authors would contribute single paper only. In the generalized form Lotka's law, referred as inverse power law by Bookstien[3] and expressed as:

$y_x = c / x^n$; where $x = 1, 2, 3, \ldots, x_{max}$,

y_x represents the probability of an author to publish "x" times,

'*c*' and '*n*' are the constant to be determined for each data set;

x_{max} represents the maximum value of productivity

n is considered as a measure of inequality in the distribution of scientific field, implying that any increase in the value of *n* is accompanied by the increase of low productive authors. Askew[4] observed that the value of exponent *(n)* may lie between −1.2 and −3.8 for the Lotka's law to be considered as inverse square law. This law has been applied on numerous data set of author productivity from different fields of science and technology over the years. Author productivity has been used as proxy to demonstrate scholarly maturity of a field of research. "Bibliometric indicators have been used to provide an inclusive perspective on the growth of the collective scholarly knowledge[5]". Lotka's law is indented to assess author productivity at aggregate levels. Comparison with Lotka's law is the most widely cited method for the assessment author productivity in any subject field[6].

Objectives of the study

1. To examine the validity of Lotka's Law, both in generalized and modified forms as suggested by Pao[7] and Nicholls[8];

2. Identifying the main elements involved in fitting Lotka's law; and

3. To establish the level of author productivity within Indian ANNs researchers and analyse the characteristics of prolific authors to establish any underlying trends.

Methodology

Potter[9] observed that most of the studies are not following Lotka's original methodology and lacked the coverage, in terms of either subjects or time, of the sources, making their comparison superfluous and leading to unreliable results at best[10]. Pao[11] took the initiative to standardise methods for the application of Lotka's law much similar to the original study. In 1986, and applied the same methodology on 48 different data sets and found that the majority of data sets were complied with Lotka's law. However, Nicholls[8] tried maximum likelihood method to estimate the parametric values of Lotka's law distribution and observed that it is better option. Huber and Wagner-Dobler[12] commented that sample size should have more than one thousand unique authors as there is high percentage of low productive authors in a specific field of research.

The present study is a secondary analysis of bibliometric data downloaded from Science Citation Index-Expanded on ANNs research in India. The data spanned over 24 year period from 1991-2014 and included original research, review articles and proceeding papers only. All the authors were included in the analysis to ensure that any comparisons are consistently using the same data set. The least squares method was used to identify the productivity gradient where the number of data entries, N, is 8, X is the logarithm of the number of articles published (1, 2, 3, …) and Y is the logarithm of the number of authors:

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2}$$

To test the applicability of Lotka's law on the set of data, Kolmogorov-Smirnov (K-S) statistic, suggested by Coile[6], is used to determine the maximum deviation D:

$$D = \mathrm{Max}\left| F(x) - S(x) \right|$$

where F(x) is the theoretical cumulative frequency function and S(x) is the observed cumulative frequency function of a sample. The K–S test requires calculation of the fraction of authors expected to publish one article (c) within the sample and also uses 'p' to represent the number of articles published:

$$C = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n+1}}}$$

Application of Lotka's law

The procedure to apply Lotka's law requires a log-log graph between number of authors and number of papers. Only those pair of point which lied on a relatively straight line was considered for calculating parameter of the Lotka's law. First eight points were observed as lying on a straight line. (Figure 1 and Table 1).

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} \quad \ldots (1)$$

$$n = \frac{8*47.4342 - (10.6045*41.9959)}{8*17.5202 - (10.6045)\text{^}2} = -2.35$$

$$C = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n+1}}} \quad \ldots (2)$$

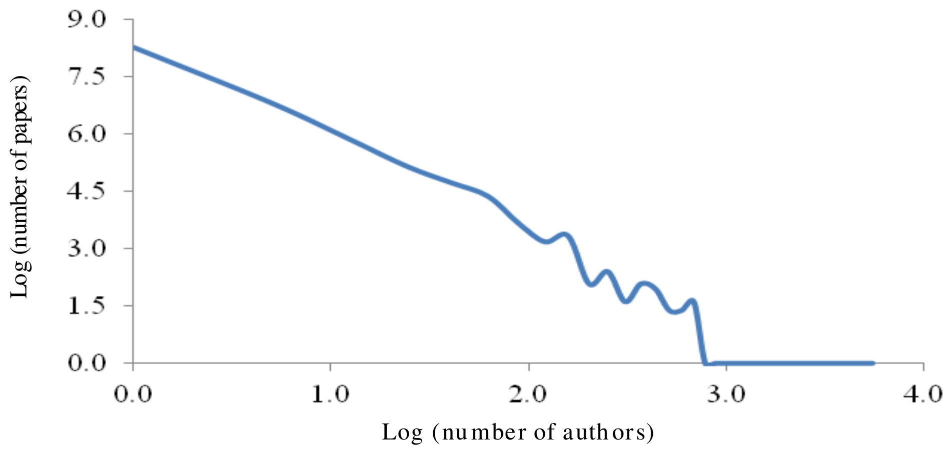

Fig. 1—Log-log plot of number of authors and number of articles for Artificial Neural Networks Research in India during 1991-2014

Table 1—Procedure to calculate parameters of Lotka's law

Number of Papers (x)	Number of Authors (y)	X=Log(x)	Y=Log(y)	XY=Log(x) * Log(y)	X^2=Log(x)^2
1	3879	0.0000	8.2633	0.0000	0.0000
2	914	0.6931	6.8178	4.7258	0.4804
3	345	1.0986	5.8435	6.4198	1.2069
4	171	1.3863	5.1417	7.1279	1.9218
5	112	1.6094	4.7185	7.5941	2.5902
6	79	1.7918	4.3694	7.8290	3.2105
7	39	1.9459	3.6636	7.1290	3.7865
8	24	2.0794	3.1781	6.6086	4.3239
Total		10.6045	41.9959	47.4342	17.5202

To calculate first part of the above equation for the constant *c,* value of n = 2.35 is substituted from equation 1 by least square method. (Value of p is chosen 20. After that point the equation gives negligible values and therefore of no use.

$\sum_1^{p-1} 1/x^n$ = First part of the equation (2).

Same way the second part of the equation=$1/(n-1)(p^{n-1})$=0.01298,

Third part of the equation=$1/2p^n$=0.0088, and the remaining forth part=$n/24(p-1)^{n+1}$=0.0000.

Summing up all four part and dividing them by 1 yields final value of the constant c=1/(1.3943+0.01298+0.0088+0.0000)=0.7062.

Table 1(a)—Different calculations of the first part of equation 2

p	Corresponding value	p	Corresponding value	p	Corresponding value	p	Corresponding value
1	1.0000	6	0.0148	11	0.0036	16	0.0015
2	0.1961	7	0.0103	12	0.0029	17	0.0013
3	0.0756	8	0.0075	13	0.0024	18	0.0011
4	0.0385	9	0.0057	14	0.0020	19	0.0010
5	0.0228	10	0.0045	15	0.0017	20	0.0009
						∑	1.3943

The total Indian research output of 3411 papers is contributed by 5654 unique authors with 0.6 authors per paper (Table 3). As immense majority 3679 (68.6%) authors contributed only one article, which is much greater number than 60% indicated by the Lotka's law. Author productivity distribution in the field of ANNs research in India was subjected to Lotka's law of author productivity using the least square method and maximum likelihood method.

Table 2 gives parameters of Lotka's law by both least square method and maximum likelihood method. Using the least squares method the exponent value of Lotka's law (n) was calculated as −2.35. As the value of n (−2.35) lies within the range of -1.2 and -3.8 implies that ANNs authorship in India follows an inverse power distribution, with most authors publishing only one article and significantly smaller numbers contributing higher publication levels. The exponent *(n)* value for ANNs research in India (−2.35) has given $c = 0.7061$, suggesting that an expected 70.62% of authors would publish one article, whereas the observed number was actually 68.61%. (Table 3). The author data can be further assessed in a comparison of observed and predicted authorship values by the K–S goodness-of-fit test. As the Dmax value was within the K-S critical value, at 10% level, suggesting that least square method was appropriate to explain the author productivity in the field of ANNs research in India.

To apply maximum likelihood method, LOTKA, a computer program by Rousseau[13], was used. Lotka's law parameters, n, c are calculated. This program provides K-S statistics as well. Author productivity distribution of ANNs research in India was submitted to the program and analysed. This program resulted in $n = 2.37$, c=0.7151, Dmax = 0.0290 and K-S critical value = 0.0217. As the Dmax is more than the K-S critical value implies that maximum likelihood method was not able to account for the author productivity distribution of ANNs research in India.

Evaluation of Prolific Authors

Most of the prolific authors listed in Table 4 are from different top rated engineering institutions in India including Indian Institute of technology(s), National Institute of technology(s) and others. There are merely 20 authors contributed more than 15 articles. The largest number of articles by one author is 42. The second and third ranked prolific authors contributed 34 and 28 respectively. This shows the extremely large number of publication contributed by a single author in ANNs research in India. There are only 9 authors who have published 20 or more papers during the period of study on the topic of ANNs.

Table 4 lists 16 most productive authors, who published more than 17 articles (about ½ a percentage of the total number of papers), and their number of publications and number of times they were first authors in their respective publications. A detail examination of the top author with 42 publications, Ganapati Panda, indicates that his first publication in ANNs research appeared in 1994 and he is still active in 2014. Whereas, author ranked second with 34 publications, published his first paper in 2007 and he is also still active in 2014. Similarly, author ranked third with 30 papers, Sudheer, K. P., started contributing in 2002 and he also is still active. This indicates that these top ranked authors are likely to contribute in the progress of ANNs research.

Most of the prolific authors were able to attract citations more than the average citations of the entire data under the study. However, it is also observed that almost all the papers contributed by the prolific authors were in collaboration with one or more authors either from within the country or from other countries. It confirms that co-authorship is responsible for higher number of citations[14]. While studying profile of the individual authors from respective institutional official websites, it was observed that the most prolific authors occupying senior most positions such as deputy director, head of department and professor etc., However, the present study has indicated that these prolific authors were listed other than first authors much more number of instances than list as first author. (Table 4). This confirms that taking all the authors into account for the analysis will give a true picture of author productivity and collaboration profile as well. Influence of

Table 2—Comparison of parameters of the Lotka's law using least square method and maximum likelihood method

Method for calculating exponent *(n)*	*n*	c	Dmax	K-S critical value
Least Square Method	-2.35	0.7062	0.0201	0.0217
Maximum Likelihood Method	-2.37	0.7151	0.0290	0.0217

Table 3—Distribution of ANNs author productivity in India during (1991-2014)

Number of Papers	Number of Authors	Observed Fraction of Authors	Cumulative Observed Fraction of Authors	Expected Fraction of Authors	Cumulative Expected Fraction of Authors	Dmax
1	3879	0.6861	0.6861	0.7062	0.7062	0.0201
2	914	0.1617	0.8478	0.1385	0.8447	0.0031
3	345	0.0610	0.9088	0.0534	0.8981	0.0107
4	171	0.0302	0.9390	0.0272	0.9253	0.0137
5	112	0.0198	0.9588	0.0161	0.9414	0.0174
6	79	0.0140	0.9728	0.0105	0.9519	0.0209
7	39	0.0069	0.9797	0.0073	0.9592	0.0205
8	24	0.0042	0.9839	0.0053	0.9645	0.0194
9	28	0.0050	0.9889	0.0040	0.9685	0.0204
10	8	0.0014	0.9903	0.0032	0.9717	0.0186
11	11	0.0019	0.9922	0.0025	0.9742	0.0180
12	5	0.0009	0.9931	0.0021	0.9763	0.0168
13	8	0.0014	0.9945	0.0017	0.9780	0.0165
14	7	0.0012	0.9957	0.0014	0.9794	0.0163
15	4	0.0007	0.9964	0.0012	0.9806	0.0158
16	4	0.0007	0.9971	0.0010	0.9816	0.0155
17	5	0.0009	0.9980	0.0009	0.9825	0.0155
18	1	0.0002	0.9982	0.0008	0.9833	0.0149
19	1	0.0002	0.9984	0.0007	0.9840	0.0144
20	1	0.0002	0.9986	0.0006	0.9846	0.0140
22	1	0.0002	0.9988	0.0005	0.9851	0.0137
23	1	0.0002	0.9990	0.0004	0.9855	0.0135
24	1	0.0002	0.9992	0.0004	0.9859	0.0133
27	1	0.0002	0.9994	0.0003	0.9862	0.0132
28	1	0.0002	0.9996	0.0003	0.9865	0.0131
30	1	0.0002	0.9998	0.0002	0.9867	0.0131
34	1	0.0002	1.0000	0.0002	0.9869	0.0131
42	1	0.0002	1.0002	0.0001	0.9870	0.0132
Total	5654					

collaboration on author productivity and higher rate of citations also has been associated with increase in h-index of an individual author[15].

Conclusion

Lotka's law is essentially to predict author productivity at an aggregate level[16]. Therefore it cannot be used for productivity of the individual authors. The present study has provided an overview of author productivity in ANNs research in India during 1991-2014. The study found that the authors productivity correlate with Lotka's law and demonstrates a pattern of productivity in line with other professions, with a significant number of one-time authors and small number of recurring author names. Egghe[17] concluded that a smaller value of the Lotka's parameters n and c are indicative of newly emerging field of research. In the present study value of n and c are on a little higher side implying that the ANNs field has relatively matured in India. It was also observed that if Pao's methodology is applied correctly on proper sample size in terms of unique number of author and coverage, it would give better results than the methodology suggested by Nicholls. The international profile of prolific authors indicated that an evolving research base and confirms that

Table 4—Prolific authors in Artificial Neural Networks Research in India (1991-2014)

Sl. no.	Prolific Authors	Number of Papers First author)	Citations Per Paper
1	Panda, Ganapati, IIT, Bhubaneswar	42(6)	14.7
2	Samui, Pijush, VIT University, Vellore	34(18)	4.2
3	Sudheer, K. P., IIT, Madras	30(8)	46.3
4	Tambe, S. S. , National Chemical Laboratory, Pune	28(0)	16.1
5	Kulkarni, D. B., Gogte Institute of Technology, Belgaum	27(7)	16.2
6	Singh, T. N., IIT, Bombay	24(5)	20.0
7	Raj, Baldev, Indira Gandhi Center for Atomic Research, Kalpakkam	23(1)	8.7
8	Deo, M. C., IIT, Bombay	22(1)	16.9
9	Pal, S. K., Indian Statistical Institute, Calcutta	20(6)	25.1
10	Kalra, P. K., IIT, Kanpur	19(0)	4.5
11	Mahapatra, S. S., National Institute of Technology, Rourkela	18(2)	10.8
12	Chattopadhyay, Surajit, Pailan College of Management & Technology, Kolkata	17(15)	7.4
13	Balaji, C., IIT, Madras	17(3)	5.2
14	Khandelwal, Manoj, Maharana Pratap University of Agriculture & Technology, Udaipur	17(9)	10.9
15	Majumdar, Abhijit, IIT, Delhi	17(9)	9.8
16	Shukla, Anupam, ABV Indian Institute of Information Technology & Management, Gwalior	17(5)	1.2
	Sub-total	372	14.8
	Total	3411	9.3

IIT = Indian Institute of Technology

collaboration increases individual productivity and visibility by enhancing citations to their publications.

References

1. Aaron L, Baker S, Gill J, Jackowski M, Johnston J, Kowalczyk N, Legg J, Leggett T, Metcalf K, Scutt D and Wilson B, Building the body of knowledge, *Radiology Technology*, 82 (2) (2010) 185–187.
2. Lotka A J, The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, 16 (12) (1926) 317-323.
3. Bookstein A, The bobliometric distribution, *The Library Quarterly: Information, Community, Policy*, 46 (4) (1976) 416-423.
4. Askew C A, *An examination of Lotka's law in the field of library and information studies,* Ph.D thesis, Florida International University, 2008. Available at: http://digital-commons.fiu.edu/etd/182 (accessed on 30 Nov 2015)
5. Coursaris C K and Osch W V, A scientometric analysis of social media research (2004–2011), *Scientometrics*, 101(1) (2014) 357–380.
6. Coile R C, Lotka's frequency distribution of scientific productivity, *Journal of the American Society for Information Science,* 28(6) (1977) 366-370.
7. Pao M L, Lotka's Law: A testing procedure, *Information Processing & Management,* 21(4) (1985) 305-320.
8. Nicholls P T, Empirical validation of Lotka's law, *Information Processing & Management,* 22(5) (1986) 417-419.
9. Potter W G, Lotka's Law Revisited, *Library Trends*, 31(2) (1981) 21-39.
10. Schorr A E, Lotka's Law and map librarianship, Journal of the Association for Information Science and Technology, 26 (3) (1975) 189-90.
11. Pao M L, An empirical examination of Lotka's Law, *Journal of the American Society for Information Science,* 37 (1) (1986) 26-33.
12. Huber J C, and Wagner-Dobler R., Scientific production: A statistical analysis of authors in mathematical logic, *Scientometrics,* 50 (2) (2001) 323-337.
13. Rousseau B and Rousseau R, Lotka a program to fit a power law distribution to observed frequency data, *CYBERmetrics*, 4 (1) (2000) paper 4.
14. Harande Y I, Author productivity and collaboration: An investigation of the relationship using the literature of technology, *Libri,* 51 (2) (2001) 124–127.
15. Figg W D, Dunn L, Liewehr D J, Steinberg S M, Thurman P W, Barrett J C and Birkinshaw J, Scientific collaborations result in higher citation rates of published articles, *Pharmacotherapy,* 26 (6) (2006)759–767.
16. Newby G B, Greenberg J and Jones P, Open source software development and Lotka's Law: Bibliometric patterns in programming, *Journal of the American Society for Information Science and Technology*, 54 (2) (2003) 169-178.
17. Egghe L, Theoretical evidence for empirical findings of A. Pulgarin on Lotka's law, *Malaysian Journal of Library & Information Science,* 17 (3) (2012) 1-5.