

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

9-21-2021

Application Of Lotka's Law To The Research Productivity In The Field Of Open Educational Resources During 2011-2020

Nagaiah M
nagaiahben@gmail.com

Dr. Thanuskodi S

Dr.Alagu A

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>

M, Nagaiah; S, Dr. Thanuskodi; and A, Dr.Alagu, "Application Of Lotka's Law To The Research Productivity In The Field Of Open Educational Resources During 2011-2020" (2021). *Library Philosophy and Practice (e-journal)*. 6365.

<https://digitalcommons.unl.edu/libphilprac/6365>

Application of Lotka's Law to the Research Productivity in the field of Open Educational Resources during 2011-2020

Nagaiah M

Research Assistant/ Research Scholar
Department Library and Information Science
Alagappa University, Karaikudi, 630 003.
E.mail: nagaiahben@gmail.com

Dr. Thanuskodi S

Head of the Department
Department Library and Information Science
Alagappa University, Karaikudi, 630 003.
Email: thanuskodi_s@yahoo.com

Dr. A. Alagu

Teaching Assistant
Department of Library & Information Science
Alagappa University, Karaikudi, 630 003.
E.mail: alagubharathilis@gmail.com

Abstract

This article reveals the scientific productivity of authors in the field of open education resources literature and the applicability of Lotka's law in the literature of open education resources for a given data set. Data was retrieved from the WoS database for this study, and a total of 5075 documents were published by 1126 authors in 2011-2020. Goodness of fit test - Kolmogorov-Smirnov (K-S) test was also used to measure the practicability of Lotka's law in the field of OER open educational resources. The productivity distribution does not fit when Lotka's law is used applied in generalized form and in its original form in the data set. This confirms that Lotka's law doesn't fit to the literature of open educational resources (OER).

Keywords: Lotka's Law, Kolmogorov – Smirnov (K-S) test, data analysis, scientific productivity, Open Educational Resources.

Introduction

This study carry out the perspective of Open Educational Resource research using Bibliometrics analytics indicators, then classifies open educational resources research with text mining techniques, and creates evolutionary trends in open educational resource research. "The

merriam-webster.com/dictionary are definition of bibliometrics is the application of statistical methods to the study of bibliographic data How many publishing scientists know anything at all about citation networks bibliometrics, or even basic information retrieval?— Donald A. Windsor, Chemical & Engineering News, 18 Nov. 1974”

The statistical analysis of the scientific literature is an important part of bibliographic measurements, which deals with the applications of bibliometrics laws. The three bibliometrics laws are the Bradfords law, Lotka’s law, and Zipf’s law. One of the lotka’s laws mainly used for the scientific productivity of authors.

In 1926, Alfred J. Lotka’s law published his pioneering study Frequency distribution of scientific productivity determined from a decade index (1907–1916) of chemical contractions. Lotka concluded:

“The number of authors making n contribution is about $1/n^2$ of those making one; and the proportion of all contributors, that makes a single contribution, is about 60 percent.”

This study attempts to applicability of the Lotka’s Law in Open Educational Resource scientific productivity that helps identify author in the literature published from 2011 to 2019. Lotka's Law (1926) stated that if the frequency distribution of scientific productivity is "possible, then different capable men will be interested in determining the area that contributes to the advancement of science by considering the first simple scale of production." Alfred J. Lotka’s; (1926).

Review of Literature

Several studies have been applied in which lotka’s law seeks to take and apply different datasets of different subject domains. Thanuskodi, & Venkatalakshmi, (2010) has described a study focused on apply bibliometric analyses the performance of research output of scientist in the university of Tamil Nadu. Its examined area of research group, research department in universities to map the intellectual structure of research. Output of 501 records that were extracted from web of science was an online resources that combine are science citation Abstract, Social Science Citation Index, and Arts & Humanities citation index. In this study covered SCIE about 5900 journals, SCI covers about 3500, the SSCI 1700 and 3300 selectively , while the A & HCI covers more than 1100 approximately 7000 journals. The studied output in a specific field of knowledge, individual subject output, and author publication.

Alagu, & Thanuskodi, S. (2019) analyzed the publication methods as a whole depends on the literature, and examined the digital literacy research productivity during the period of 1992-2011. The bibliographic data extracted from the web of science documents include 512 publications, and this data was analyzed using the Histcite software application. The most productive countries and the largest number of articles have been published in the form of total articles covering 512 entries. Maximum number of documents 126 Published in 2011. Adolescent and Adult Literacy Magazine is ranked first with 18 records.

Kumar & Senthilkumar (2019) applied Lotka's law in Astronomy & Astrophysics literature for a given data set to express the scientific productivity pattern of authors from India in the field of Astronomy & Astrophysics research. Data for the study was retrieved from WoS database which comprises a total of 6363 papers published by 2,719 authors during 2013-2017 in straight count of the authorship. Applicability of Lotka's Law was examined by Lotka's Inverse square method and general power method were used along with Chi-Square test and Kolmogorov-Smirnov (K-S) test were also used to test the viability of Lotka's Law in the same field of study. From this study they concluded that the productivity distribution didn't when Lotka's was applied to the given dataset and it lead to the confirmation that Lotka's law doesn't fit to the literature of Astronomy & Astrophysics.

Kherde (2020) Lotka's law of scientific productivity is attempted to apply in the domain of DESIDOC Journal of Library & Information Technology. Articles published since its first to 38th volume is taken for the study by author. Kolmogorov-Smirnov test (K-S) test is applied for the verification of the dataset. Using Lotka's law number of authors published single article is identified. In the same way contribution by the authors in publishing two, three, four and so on articles are identified. On analysing the data author concluded that the law does not fit to the collection of dataset.

SB & Batcha (2020) applied Lotka's law of the author's productivity in the field Thermodynamic between the year 2015 and 2019. The Goodness of fit test Kolmogorov-Smirnov (K-S) test was applied to the dataset of research in thermodynamics. The dataset for the study is derived from the 'Web of Science' in the form of plain text using the keyword 'Thermodynamics' OR 'Basics of heat and temperature'. From the findings they concluded that, in the field of thermodynamics, significance of Lotka's law is 0.003 implying that, Lotka's law is not applicable for this dataset. This study then identified the most productive author and journal from the dataset.

Ivanova, Grosseck, & Holotescu, C. (2020) In their study, the current research level, features of OER, tools used for the development of OERs, and analysis and learning analysis contributes to the formation, development and reuse of OER. Study methods was analyzed and used on dataset extraction from abstract and citation database Scopus, web of science and dimensions, bibliographic data processing and visualization of bibliometric built with VOS viewer and biblioshiny. The approach proposed in this study is also applied to the development of predictive models based on supervised machine learning methods such as linear regression and artificial neural networks. They are capable of predicting the future growth of scientific production over the area of OERs.

Rathika, Thanuskodi, & Sudhakar, (2020) studied of Lotka's law and the scientific productivity of marine pollution research was conducted with the aim of evaluating the development of marine pollution literature with a scientific approach published since 1989 and coded 2018 and Scopus multidisciplinary online database. Compliance with the law of the Lotka's tested with a Chi-square test to evaluate what was expected values and observed Values. The calculated Chi -square value (1787.57) was greater than the Chi-square value, which is 47.40 33 at the level of significance is 0.05. It was confirmed that the Lotka's law does not apply

to the literature on Marine Pollution. There are 2219 source headlines/magazines it also contributed to marine pollution research. The Marine Pollution bulletin issued by Elsevier Ltd. of England Top with 2961 (20.98%) publications. It was followed by the 2005 International Oil Spill Conference, IOOSC 2005 Published by American Petroleum Company, and the USA with 662 (4.69%) releases. Overall Growth Rate Calculated Various scientific methods in the field of marine pollution have been growing significantly internationally in recent years. On the other hand, the growth rate of marine pollution in India was moderate.

Objective of study

1. Exploring the scientific productivity of teachers in the field of open education resource literature.
2. To verify the Lotka's law in the field of open education resources in its general form and in Inverse Square.
3. To examine the validity of Lotka's law using the method of direct counting by the authors
4. To apply the non- parametric Kolmogorov-Smirnov (K-S) test, which is a 'Goodness-of-fit' parameter in accordance with Lotka's law.

Methodology and data sources

In this study, we followed Pao (1985) discussion on the procedure to measure Lotka's Law .The data for this study was obtained in plain text format from 'Web of Science'. The term 'open education resources' is used to find data from the web of science database.'Topic' is metadata category used to retrieve data. Customize by year-month-date option; in this study selected for the period 2011-01-01 to 2020-12-31. A total of 1126 records are being analyzed. Data is analyzed and calculated with the help of histcite, and Microsoft Excel. The indicators used for data analysis and the Lotka's law are as follows. The lotka's formula for determine validity was conducted by the K-S goodness of fit test. The K-S goodness-of-fit test was conducted to determine the validity of Lotka's formula, when predicting publication productivity in "Open educational Resources".

Null Hypothesis:

The null hypothesis represents the significant difference between the observed distribution and the theoretical distribution predicted by Lotka's law, and difference between the critical value and deviation value valid test contacted by K-S goodness of fit test.

- Lotka's inverse Square Law of scientific productivity fits OER selected dataset.

Lotka's law:-

As an empirical law that describes the distribution of authors and papers, Lotka's law can be explained from various aspects. In the field of infometrics, Lotka's law refers to the distribution of authors over a period of time or within certain subject areas. From the perspective of anthropology, Lotka's law can be understood in the following statement, in the process of human

development, and the contributions to human progress by different individuals with different personalities...” Merton considered such differences the Matthew’s effect.

Theoretical Framework

Lotka's law can be represent in the following (equation no I).

$$x^n * y = c \text{ ----- (1)}$$

X is the number of articles published (1, 2, 3...)

‘Y’ is the number of authors with frequency ‘x’ number of articles

‘n’ is an exponent that is constant for a given set of data ‘c’ a constant

When n=2, exponent n is used for the data set it’s called ‘inverse square law of scientific productivity’. In this case, the value of ‘C would be a constant 0.06079’. The value of ‘n’ differs for different data sets. To calculate the f ‘n’ value in the study mentioned about, linear least square (LLS) method has been used. This method was defined by pao (1985). This can be calculated as. (Equation no II)

$$N = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2} \text{ ----- (2)}$$

Where, ‘x’ is the number of publications; logarithm value, ‘y’ is the number of authors; its logarithm of ‘y’ value, and N’ is the number of data pairs available for data analysis. The constant ‘C’ this can be calculated by the following equation - III

$$C = \frac{1}{\sum 1/x^n} \text{ ----- (3)}$$

To examine the observed frequency pattern of the author's productivity suits the expected frequency pattern. ‘Pao’ advised applying the non-parametric Kolmogorov – Smirnov (K-S) goodness -of- fit test. To check it, the highest deviation between the observed cumulative relative frequency and expected cumulative relation frequencies needs to be considered and then compared it with the critical value (C.V.) which can be calculated by the following equation determined by Nicholls. (Equation no IV).

$$C.V = \frac{n}{\sqrt{\sum y}} \text{ ----- (4)}$$

Goodness-of Fit

There are several goodness-of fit tests are available compatibility and validity of data. The goodness –of fit tools are the most frequently applied Kolmogorov – Smirnov (K-S) test and Chi Square test in Lotka’s law. In this study, the K-S test is used to verify the law of the Lotka’s law.

K-S test

The theoretical and observed cumulative frequency distribution of the authors at each level of cumulative frequency distribution is calculated by the K-S test. Its compare with the critical value calculated using equation –IV with maximum difference value in observed value. If the difference is less than the critical value, it is accepted otherwise null hypothesis rejected.

Data Analysis:

This study describes the data analysis process used to solve the hypothesis. The procedure defined by pao (1985) has been followed to determine the extent to which lotka's law can be applied to predict productivity in the field of open education resources.

Table 1: Frequency distribution of Research contributions

X (Publication)	Y (Authors)	Percentage of Y	Total Contribution	Total Contribution Cumulative
1	197	17.49556	197	197
2	264	23.44583	528	725
3	220	19.53819	660	1385
4	147	13.05506	588	1973
5	95	8.436945	475	2448
6	58	5.150977	348	2796
7	48	4.262877	336	3132
8	27	2.397869	216	3348
9	17	1.509769	153	3501
10	10	0.888099	100	3601
11	8	0.71048	88	3689
12	10	0.888099	120	3809
13	3	0.26643	39	3848
14	3	0.26643	42	3890
15	4	0.35524	60	3950
16	1	0.08881	16	3966
18	1	0.08881	18	3984
19	1	0.08881	19	4003
20	1	0.08881	20	4023
21	1	0.08881	21	4044
22	2	0.17762	44	4088
24	1	0.08881	24	4112
25	1	0.08881	25	4137
26	2	0.17762	52	4189
28	1	0.08881	28	4217

32	1	0.08881	32	4249
38	1	0.08881	38	4287
788	1	0.08881	788	5075
	1126		5075	

Table 1 shows the frequency distribution of authors' productivity in the field of open educational resource. Of the 6363 papers has been published by, **5075** authors, 197(17.49%) produced one article, 264 (23.45%) produced two articles, 220 (19.54%) produced three articles and so on.

Application of Lotka's law:-

To determine the value of 'N', 'c' and c.v.:- The Lotka's law deals with the distribution frequency of authors productivity in a particular subject. In the present study, an attempt is being made to study the applicability of Lotka's law in the field of open educational resources. The following table is generated from data extracted from the Web of Science.

Table 2: Determining the value of 'n', 'C' and C.V

x	y	x log	Y log	xy log	x ²	x2	x ⁿ	1/x ⁿ
1	197	0	5.2832	0	0	0	1	1
2	264	0.69315	5.57595	3.86495	0.48045	0.48045	2.57398	0.3885
3	220	1.09861	5.39363	5.92551	1.20695	1.20695	4.475	0.22346
4	147	1.38629	4.99043	6.91821	1.92181	1.92181	6.62537	0.15094
5	95	1.60944	4.55388	7.32918	2.59029	2.59029	8.98246	0.11133
6	58	1.79176	4.06044	7.27534	3.2104	3.2104	11.5186	0.08682
7	48	1.94591	3.8712	7.53301	3.78657	3.78657	14.2139	0.07035
8	27	2.07944	3.29584	6.8535	4.32408	4.32408	17.0535	0.05864
9	17	2.19722	2.83321	6.22521	4.8278	4.8278	20.0257	0.04994
10	10	2.30259	2.30259	5.3019	5.3019	5.3019	23.1206	0.04325
11	8	2.3979	2.07944	4.98628	5.7499	5.7499	26.3305	0.03798
12	10	2.48491	2.30259	5.72171	6.17476	6.17476	29.6485	0.03373
13	3	2.56495	1.09861	2.81788	6.57897	6.57897	33.0688	0.03024
14	3	2.63906	1.09861	2.8993	6.96462	6.96462	36.5863	0.02733
15	4	2.70805	1.38629	3.75415	7.33354	7.33354	40.1965	0.02488
16	1	2.77259	0	0	7.68725	7.68725	43.8955	0.02278
18	1	2.89037	0	0	8.35425	8.35425	51.5456	0.0194
19	1	2.94444	0	0	8.66972	8.66972	55.4907	0.01802
20	1	2.99573	0	0	8.97441	8.97441	59.5121	0.0168
21	1	3.04452	0	0	9.26912	9.26912	63.6073	0.01572
22	2	3.09104	0.69315	2.14255	9.55454	9.55454	67.7742	0.01475

24	1	3.17805	0	0	10.1	10.1	76.3147	0.0131
25	1	3.21888	0	0	10.3612	10.3612	80.6845	0.01239
26	2	3.2581	0.69315	2.25834	10.6152	10.6152	85.1184	0.01175
28	1	3.3322	0	0	11.1036	11.1036	94.1724	0.01062
32	1	3.46574	0	0	12.0113	12.0113	112.986	0.00885
38	1	3.63759	0	0	13.232	13.232	142.832	0.007
788	1	6.6695	0	0	44.4822	44.4822	8930.21	0.00011
	1126	72.398	51.5122	81.807	224.867	224.867	10139.6	2.5087

$$\frac{N\sum xy - \sum x \sum y}{N\sum x^2 - (\sum x)^2}$$

N= number of sets of data , x = logarithm of x, (i. e. number of publications) Y = logarithm of (y, i. e. number of authors)

$$= \frac{28(81.80702) - (72.39802)(51.51221)}{28(224.8669) - (72.39802)^2}$$

$$= \frac{2290.596556 - 3729.382}{6296.2732 - 5241.4732999204}$$

$$= \frac{2290.596556 - 3729.382}{1054.7999}$$

$$= \frac{-1438.785444}{1054.7999}$$

$$\mathbf{N = -1.364}$$

$$\mathbf{C.V = \frac{n}{\sqrt{\sum y}}}$$

$$= \frac{-1.364}{\sqrt{1126}}$$

$$= \frac{-1.364}{33.55592}$$

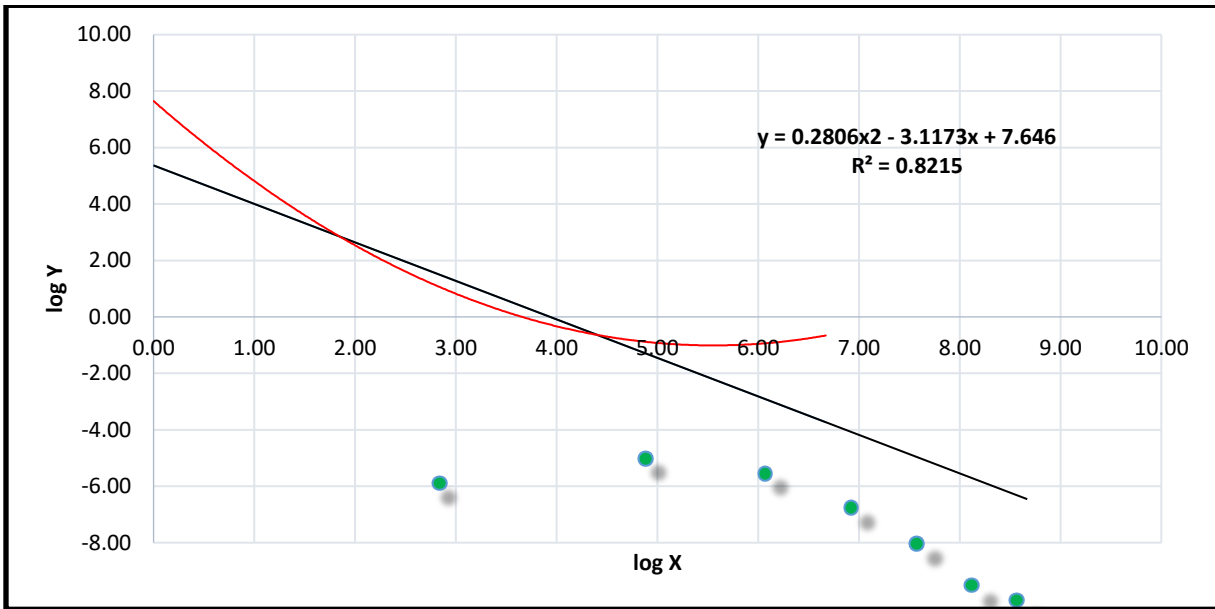
$$\mathbf{C.v = -0.0406}$$

$$\mathbf{C = \frac{1}{\sum 1/x^n}}$$

$$= \frac{1}{2.508695}$$

$$\mathbf{c = 0.39862}$$

From the above mathematical calculation n=-**1.364**, c=0.39862, c.v. = **-0.0406**



In this a plot graph where the frequency distribution categories are plotted. X is the number of papers and y is the number of authors publishing x papers. In compare to any common visual used to report on Lotka's Law, this visualization summarizes the data based on groups and add the element of colors to differentiate between the groups. In this case, we used ggplot2 to generate this bar plot.

Table: 3 Kolmogorov – Smirnov (K-S) Test (n= -1.364, C= 0.39862, C.V= -0.0406)

Publication x	No. Authors y	Observed Value		Expected Value		Deviation $\sum fo(yx) - \sum fe(yx)$
		Frequency	Cum. fo	Frequency	Cum.fe	
		$fo(yx) = yx / \sum yx$	$\sum fo(yx)$	$fe(yx) = C(1/xn)$	$\sum fe(yx)$	
1	197	0.174956	0.174956	0.6234	0.6234	-0.4484444
2	264	0.234458	0.409414	0.155782814	0.779182814	-0.369769
3	220	0.195382	0.604796	0.069236806	0.84841962	-0.2436239
4	147	0.130551	0.735346	0.038945704	0.887365324	-0.152019
5	95	0.084369	0.819716	0.02492525	0.912290574	-0.0925748
6	58	0.05151	0.871226	0.017309202	0.929599776	-0.0583742
7	48	0.042629	0.913854	0.012716964	0.94231674	-0.0284624
8	27	0.023979	0.937833	0.009736426	0.952053166	-0.0142201
9	17	0.015098	0.952931	0.007692978	0.959746144	-0.0068154
10	10	0.008881	0.961812	0.006231313	0.965977457	-0.0041657
11	8	0.007105	0.968917	0.005149845	0.971127302	-0.0022108
12	10	0.008881	0.977798	0.0043273	0.975454603	0.0023429

13	3	0.002664	0.980462	0.003687167	0.97914177	0.00132
14	3	0.002664	0.983126	0.003179241	0.982321011	0.0008051
15	4	0.003552	0.986679	0.002769472	0.985090483	0.001588
16	1	0.000888	0.987567	0.002434106	0.98752459	4.202E-05
18	1	0.000888	0.988455	0.001923245	0.989447834	-0.0009931
19	1	0.000888	0.989343	0.001726125	0.99117396	-0.0018312
20	1	0.000888	0.990231	0.001557828	0.992731788	-0.0025009
21	1	0.000888	0.991119	0.001412996	0.994144784	-0.0030258
22	2	0.001776	0.992895	0.001287461	0.995432245	-0.002537
24	1	0.000888	0.993783	0.001081825	0.99651407	-0.0027308
25	1	0.000888	0.994671	0.00099701	0.99751108	-0.0028397
26	2	0.001776	0.996448	0.000921792	0.998432872	-0.0019853
28	1	0.000888	0.997336	0.00079481	0.999227682	-0.001892
32	1	0.000888	0.998224	0.000608527	0.999836209	-0.0016124
38	1	0.000888	0.999112	0.000431531	1.00026774	-0.0011558
788	1	0.000888	1	1.00352E-06	1.000268744	-0.0002687
	1126					0

$$D = \text{Max} | F_o(X) - S_n(X) |$$

Where $F_o(X)$ is the cumulative frequency of expected value

$S_n(X)$ is the cumulative frequency of observed value.

The K-S goodness – of –fit test is used to test the validity of Lotka's law on the observed frequency of authors compared to the expected frequency of authors' productivity. Looking at the thirteen columns in Table 3, the maximum deviation between the overall distributions was found to be $D_{\text{max}} = 0.00132$ which is higher than the observed significant value, i.e. C.V. = -0.0406. Therefore, it can be confirmed that the K-S test reveals that the current data set does not confirm the applicability of Lotka's law in the field of open education resources (OER) research.

Conclusion

Lotka's law can scientifically solve the distribution of authors in scientific bibliometric. However, some problems arise in the scientific application of Lotka's law. In 1926 Alfred Lotka had put the law regarding the scientific productivity on the basis of the datasets taken from WoS. Lotka's law of the author's productivity is considered one of the classical laws of the bibliometric mathematically, this law can be expressed as $a(n) = a(1)/n^2$. In this study, we followed Pao's modified methods, determine the goodness of fit with a set of observed values and value which is predicted theoretically. K-S test has been applied to verify the applicability of the Lotka's Law on the data set. The K-S test has that at the 0.01 level of significance the maximum difference of **0.00132** at the critical value of (**-0.0406**), which is higher than the critical value. Therefore, the K-S tests rejected the null hypothesis these data not fit modified Lotka's law. To conclude from

the studied data set that Lotka's Law of scientific productivity rules doesn't fit to the open educational resources research literature.

References

- Alagu,A.,&Thanuskodi,S.(2019).Bibliometric analysis of digital Literacy research output: A global perspective. *Library Philosophy and Practice*,1-19.
- Ayyanar,K.,&Alagu,A.(2020).Growth and Authorship Pattern of Norovirus Research Output: A Scientometric Study. *Handbook of Metric Studies for Library and Information Science Schplars*, 70-81.
- Friedman,A(2015). The power of Lotka's law through the eyes of R.Romanian *Statistical Review*,2(63),69-77
- Ivanova,M.,Grosseck,G.,&holotescu, C.(2020).Analyis and Modeling the Domain of Open Educational Resources From Learning Analytics Perspective. *eLearning &Software for Education*, 3.
- Kherde, M. R. (2020). Applicability of Inverse Square Law of Scientific Productivity in DESIDOC *Journal of Library & Information Technology. Library Philosophy and Practice*, 1-8.
- Kumar, S., & Senthilkumar, R. (2019). Applicability of Lotka's law in astronomy & astrophysics research of India. *Library Philosophy and Practice*, 1-13.
- Merriam- Webster.(n.d.).Bibliometrics.In Merriam – Webster.com dictionary. Retrieved July 13,2021,from <https://www.merriam-webster.com/dictionary/bibliometrics>.
- Naheem,K.T.,Sivaraman,P.,&Saravanan,G.(2019).Application of Lotka's Law in Parasitology research output of India.*Library Philosophy and Practice*, NA-NA.
- Nicholls,P.T.(1985)Empirical validation of Lotka's law .*Information Processing & Management*,22(5),417-419.
- Nicholls,P.T.(1986)Bibliometric Modeling processes and empirical validaty of Lotka's law .*Journal of American Society for Information Science*, 40(6),379-385.
- Pao,M.L.(1985)Lotka's Law: a testing Procedure. In: *Information processing & Management*,21(4),305-320
- Rathika, N., Thanuskodi,S.,& Sundhakar,K.(2020).Lotka's law and the pattern of scientific productivity in the marine pollution research. *International Journal on Emerging Technologies*,11(2),332-341.
- SB, C., & Batcha M, S. (2020). Application of Lotka's Law to the research productivity in the field of Thermodynamics during 2015-2019.

Thanuskodi,S.,&Venkatalakshmi,V.(2010).The Growth and development of research on ecology in India : a Bibliometric study. Library Philosophy and practice,1.

Venkatesan,M.n.,&Thanuskodi,S.(2014).A Scientometric analysis of nuclear power generation research: Study. International Journal of Library and Information Studies,4(3),65-75.

Acknowledgment

This article has been written with the financial support of UGC STRIDE
Component-I grant sanctioned vide Letter No. F. 2-5/2019(STRIDE-I)
Dt.03.12.2019