A NEW WAY OF DATA VISUALIZATION USING BRADFORD'S AND LOTKA'S LAW

Akhilesh, Ajith Kumar.,* Sahoo, Subhranshubhusan, ** Hilala, Ariyana,* Sinha, Prashant Kumar. ******

ABSTRACT

{Metrics has been studied and applied in almost every research fields, such as Library and Information Science (LIS), Sociology, history of science, business, management and research policy etc. The three important empirical laws in Bibliometrics are Bradfords, Zipf's and Lotka's law are based on some assumptions. These laws gives two-dimensional view of distributions. In this paper our attempt is to give a new perception of data visualisation, which in future can be used as a tool in promoting research and developent in field of scientometrics. We can apply visualizational tools like R and Python. Finally, the paper concludes by visualizing the features from the emprical laws in 3D plotting}

Key words: *Bibliometrics, Bradford's law , LOTKA'S LAW Zipf Law, Data Visualization*

1.INTRODUCTION

Bibliometrics is the branch of library science concerned with the application of mathematical and statistical analysis to bibliography i.e. the

*Akhilesh,A.k. * akhileshajithkumar@gmail.com,*
*Sahoo,S** subhranshubhusan@gmail.com,*
*Hilala, Ariyana *** hills.ariyana@gmail.com,*
*Sinha, Prashant Kumar **** pk99597@gmail.com : (MSLIS-2nd Year, DRTC Indian Statistical Institute, Bengaluru Karnataka:)*

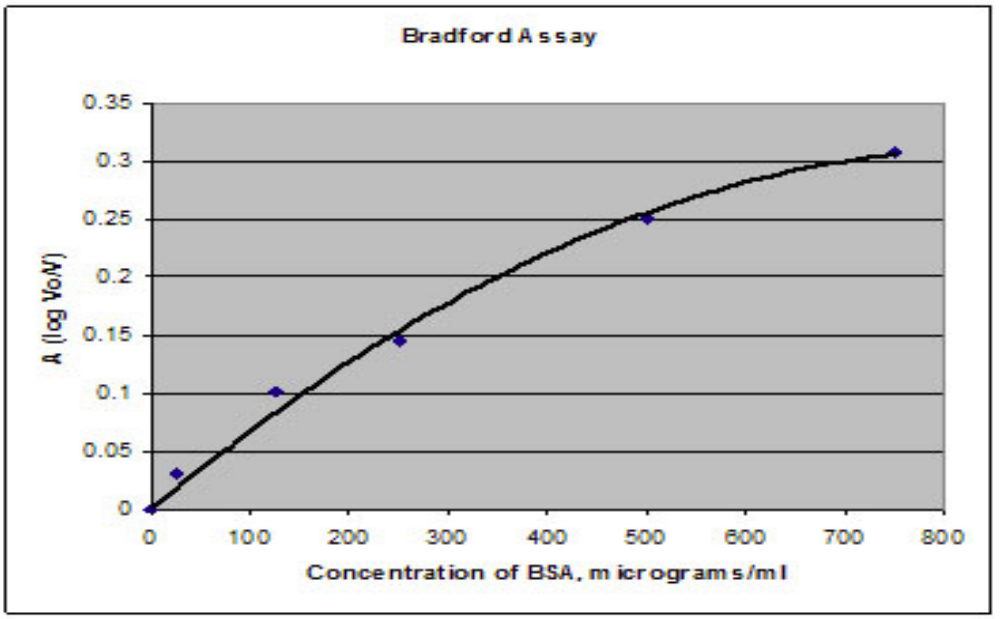
statistical analysis of books, articles, or other publications. In other words data about publications, or citation frequency. [1] In recent times, bibliometrics has caught the attention of many individuals in the field of Library and Information science. People have realized that application mathematical and statistical analytical tools on the bibliographic information gives us the oppurtunity of providing a theory that may resolve many practical problems.

It is claimed that the patterns of authors productivity, literature growth rates and related statistical distributions can be used to evaluate authors, journals, assess disciplines, and manage collections. Yet, it is unclear that a particular method could solve or test the theory and give its appropriate explanation or its predictive phenomenon.The Bradford, Lotka and Zipf distributions are considered the basic laws of bibliometrics, and each of these distributions was empirically derived. Hence these are also called as empirical laws. The distributions in each of these laws are similar to each other as special cases of a hyperbolic distribution.

Bradford Law states that: Documents on a given "subject" is distributed (scattered) according to a certain mathematical function so that a growth in papers on a subject requires a growth in the number of journals/information sources. The numbers of the groups of journals to produce nearly equal numbers of articles is roughly in proportion .

1: n: n2

Where n is called the Bradford multiplier.Bradford's law states that a small core of, for example, journals have as many papers on a given subject as a much larger number of journals, n, which again has as many papers on the subject as n2 journals.



\.

Fig. 1

(Source:http://lightboxkit.com/index_files/Images/Bradford%20Assay%20Graph.jpg)

Lotka's Law states that: The number of authors making n contribution is about $1/n^2$ of those making one contribution, and the proportion of all contributors who make a single contribution is about 60 percent and proportion of authors making two contributions is $1/2^2 x60 = 15\%$

$$a(n) = k/n^2$$

Where a is the number of authors producing n papers and k a constant.

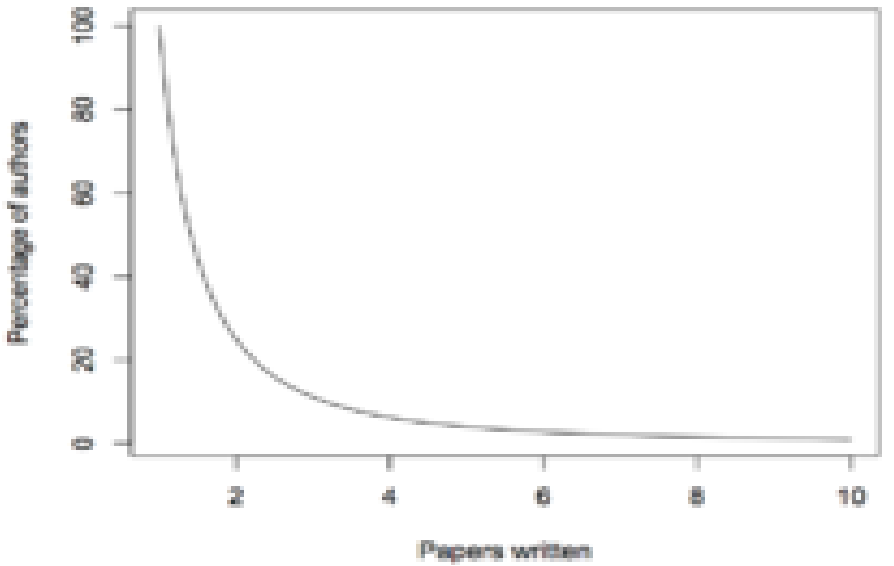


Fig. 2

Source:***https://upload.wikimedia.org/wikipedia/commons/thumb/1/1e/Lotka_plot.png/220px-Lotka_plot.png***

Zipf Law states that: If words are ranked on the basis of the frequency of their occurrence in a long text in decreasing order, the frequency of occurrence f(r) of a word is the reciprocal of its rank n (Zipf, 1949).

$$rf = c$$

r is Rank of a word, f is frequency.

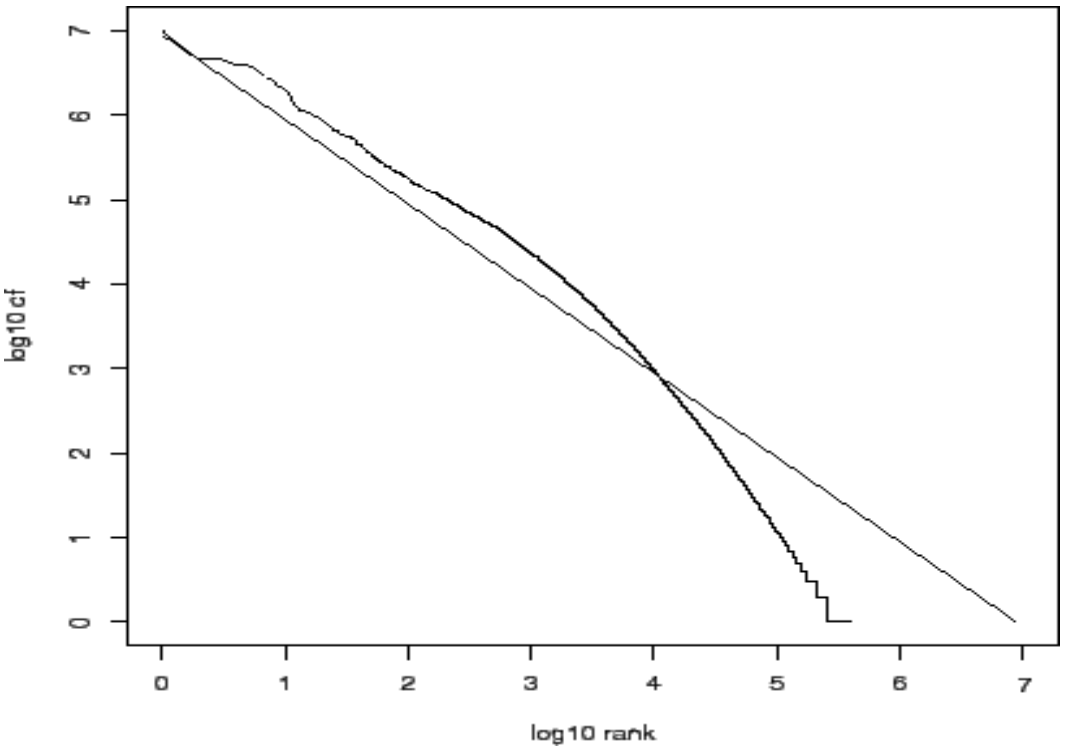


Fig. 3

Source:http://nlp.stanford.edu/IR-book/html/htmledition/img253.png

But in this paper we have taken Bradford's law and Lotka's law,then extracted 3 features and plotted in three dimensional graph(3D) to give a different view of the existing empirical laws.

2. OBJECTIVE

The main objective of the paper is to review the bibliometric Laws and on the basis of our understanding we propose a methodology consist of a new perception in viewing the current data scenarios using tools like

R , Python programming languages by plotting a 3D graph in between the features like article, author and journal extracted from the Laws.

3. METHODOLOGY

We the professionals in LIS have used Lotka's and Bradford's law to analyze the pattern in the publications which pertain to the subject and to determine the growth patterns in journals pertaining to a subject.

But these laws are generally based on assumptions and too many constrains, but here we are giving you a second view ,of how one can view the pattern in one graph and without any assumptions. We here also try to visualize the features in the above mentioned laws in a single graph. Now, here the features are authors, articles and journals which we have shown in fig(3).

Here you can use any programming language to plot the 3D graph. We here have used R programming language to plot the graph. The time period of data can be at any level, ranging from 1,5,10,20 etc. It will depend up to the granularity of the survey or research we seek to determine. We need to construct a database that will have the above mentioned features as fields, which in turn will act as the source, for future reference.

```
protype<-
matrix(c(1,35,24,2,42,36,3,17,14,4,3,3,5,2,2),ncol=3,byrow=TRUE)
> protype
     [,1] [,2] [,3]
[1,]    1   35   24
[2,]    2   42   36
[3,]    3   17   14
[4,]    4    3    3
[5,]    5    2    2
```

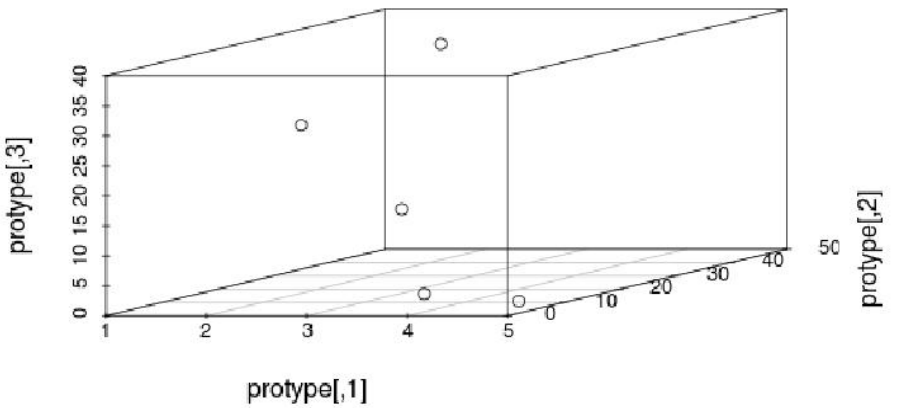


Fig. 4

4. DISCUSSION

Here we trying to look in an unconventional way how data are behaving without any interference and any predetermined assumptions of how the so called behaviour of the curve or data will pattern out. This is very important as we have to understand in contemporary times the old predictive methodologies lack to understand behavioural pattern of bibliographic data.

From the figure-4 we can identify the peak point which will inturn determine where the maximum contributions, and the publicational growth in the domain or subject has been occurred in granular level. This if we analyze over the period of time and plot out different such graphs we can find out how the peak values in each year is differing out.

From the so called peak values attained we can parallely find out the second or subsequent peak values which can be used for any analysis in a collective way or in an individual way keeping the time aspect alive in the research.

This process of determination of peaks accross years will inturn be able to help us in generating an automated process to retrieve out datas like publicational trends, collaborative trends and etc.

5. CONCLUSION

In this paper we have proposed a new approach to analyze bibliographic data in a more visualizational aspect. Thus providing a 3 dimensional overview of data representation and also discussing out the future scopes of such methodologies to help creating automated systems for analytical purposes

REFERENCES

Brookes, B. C. (1977). Theory of the Bradford law. *Journal of Documentation*, *33*(3), 180-209.

De Bellis, N. (2009). Bibliometrics and citation analysis: from the science citation index to cybermetrics. Scarecrow Press.

Li, W. (2002). Zipf's law everywhere. *Glottometrics*, *5*, 14-21.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, 46(5), 323-351.

Rao, I. K. R. (1983). Quantitative methods for library and information science. Taylor & Francis.

.

Rousseau, B., & Rousseau, R. (2000). LOTKA: A program to fit a power law distribution to observed frequency data. *Cybermetrics*, 4(1), 1-6.

Sen,B.K.(2010).Lotka's Law:a viewpoint.*Annals of Library and Information Studies*,57,166-167

University Library System. (n.p.). *University of Pittsburgh*. (Retrieved 2 September, 2015), from http://pitt.libguides.com/bibliometrics.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., ...& Venables, B. (2009). gplots: Various R programming tools for plotting data. *R package version*, *2*(4).

****