DEVELOPMENT AND ANALYSIS OF BIBLIOMETRIC MEASUREMENT TOOLS

Priya[1], Kirti Bhatia[2], Rohini Sharma[3], Shalini Bhadola[4]

[1]Student, Sat Kabir Institute of Technology and Management, Haryana, India,

[2]Assistant Professor, Sat Kabir Institute of Technology and Management, Haryana, India,

[3]Assistant professor, Government College for Women, Rohtak, India,

[4]Assistant Professor, Sat Kabir Institute of Technology and Management, Haryana, India.

Abstract: The Bibliometric is a measurable analysis of published research work in several books, journals and several printed and on line journals. It may be either informative like counting the number of articles published by the organization or estimation like assessing the influence of published research through citation analysis. Computing the number of publications is essential for few evaluations. The citation analysis permits a person to assess the effect of the articles on the work of other scholars by counting the number of times that article has been cited. Citation analysis is also used to evaluate the impact of journals, societies, and nations in several areas of research. The Institute for Scientific Information (ISI) has been a frontrunner in the field of citation analysis since a long time and publishes the Science Citation Index. In this research work we have developed citation measurement and Bibliometric indices to analyze the effect of a scholars' research work.

Keywords: Bibliometric, Citations Analysis Tool, Impact Factor, H-Index, GINI's Coefficient

Introduction: Citation analysis describes the connection between an article and the articles it has cited. It is assumed that based that there is some relation between the articles citing each other. Associations considered by citation analysis contain area, scholars, societies, and nations. Citation analysis is feasible through citation indices so that information can be linked together. The unique concept of citation indexing is accredited to Shepherd's Citations, which was initially published in 1873 as a catalog to law fiction; Eugene Garfield, creator of ISI, is accredited with starting this concept and relating it to the science innovations. Nearly all the articles, proceedings, analyses, improvements, and letters printed in scientific journals cover citations. These are cited by heading, writer, volume, issue and the name of agencies (Journal, magazine etc.) where the document was published. This supports, provides example for others, demonstrate, or intricate about the concept of author. Citations are the recognized and clear association among articles that have certain common facts. A citation index has been constructed around these associations. It catalogues articles which have been cited and recognizes the resource of the citations. Anybody organizing a literature pursuit may find any number of articles on a subject just by calculating the number of citations. Each article that is discovered gives a list of fresh citations with which to endure the search [1].

While using citation analysis, different warnings should be considered. There has been a affinity from a very long time to realize the significance of citations. The citation counts are not methods of excellence. If the measure is employed as a symbol of the excellence of someone's work without investigating the sense of the citations of that work, mainly the sort of citation and the features of those who have cited this particular work, the probability upsurges of depicting unsuitable implications regarding the influence or excellence that a person's academic and technical research has had on his subject or area [2]. For instance, there are several publication and citation features in the dissimilar arenas of science and is a requirement for scrubbing and confederacy of writers names, and in specific affiliation of writers and improvements for self-citations and self-citations. The motive for the citation also desires to be considered by the scholar. Occasionally a citation to somebody's work is not at all an authorization. It can be a severe criticism of the practice or assumptions.

The journal impact factor is one more mechanism that is consumed in Bibliometric and it has been published by IS1 in a yearly statement known as Journal Citation Reports (JCR). The impact factor is computed by dividing the total present citations of a journal obtains for papers published in the two preceding years by the total papers printed by that journal in those similar years. An impact factor of one describe that approximately each paper published in any journal round the two-year era was cited once. Clearly, the more is the impact factor, the larger will be the frequency of that journal's paper citation.

By way of with citation analysis, there is ample argument about how impact factors should and should not be exercised. Investigators are arguably afraid that strategy and subsidy conclusions can influence the usages of impact factor by scholars who do not comprehend how those scores are computed. The impact factor must not be utilized to equate journals of dissimilar areas as writers in one area can cite less paper in their articles than those in other area. Overall, major and core subject fields have advanced average impact factors than particular or functional ones. The difference is so noteworthy that the topmost journal in one area can have an impact factor lesser than the bottommost journals of other area. The category of journal creates a variance and those journals with primarily review editorials will be cited further. Due to these identical warnings, IS1 and others also advise anti-usages of the impact factor to assess the effort of specific authors.

Related Work: Authors in [3] have discussed a variety of Bibliometric indices. Yinian Gu in [4] have discussed that the current investigations symbolizes the vigorous publication movement of universal knowledge management (KM) through data composed by a discovery limited to editorial in Web of science of the ISI. A group of around 2500 distinctive scholars had funded approximately 1400 articles since 1970. The vast mainstream of them has just written one or few articles. The fruitful scholars and their work and writing rank are registered to designate their efficiency and grade of contribution in their research field. The quantity of research outcome of the first or main writers from certain countries attains 55% of the compete production. The circulation of editorials is rather extensive - they have printed around 460 headings of series, crossing over 100

JCR area classifications. The advanced eminence journals mark publication of discoveries more noticeable. A Pearson's correlation coefficient has been significant among citation occurrence of paper and impact factor of periodical, rather than of source design. The outcomes signify that cost of research and development were essentially not relative to research efficiency or citation reckonings. Since the topic extremely interrelates with other fields, the area of KM has not yet established its individual society of articles. KM may be developing an inter-punitive philosophy that is emerging at the borders of technical categories.

Author in [5] have analyzed that groups are progressively reliant on group of peoples and their perceptibility for invention. Impact of Bibliometric may be viewed as a degree of a society's perceptibility in information generating networks and can describe the reason of corporations publishing their inventions. Though, this perceptibility fluctuates through different fields. The authors have examined periodicals created by Danish companies in 1996, 1998 and 2000 to illustrate the relation of citation and association outlines in dissimilar fields. The key discoveries are in areas categorized by intercontinental association and various writers per article, intercontinental association yields in a larger amount of citations. On the other hand, nationwide association does not make any significant variance to citation influence in manufacturing research. In areas where international association and multi-writer ship is not common, an obvious figure of impact patterns cannot be achieved. Through expansion, the presented work can deliver information on usages of citations of articles in scientific journals by means of a possible space to scientific networks for societies.

Authors in [6] have discussed the problems of Citation Analysis. To conclude effects on the invention of a scientific paper, the matter of the paper should be considered thoroughly. The authors have investigated the papers in biogeography and discovered that maximum impact has not been cited and exact kinds of editorials that are significant are cited whereas other kinds of articles which are not effective have not been cited. The research that is never cited or rarely cited is being used widely. Consequently, emulative citation analysis must consider non cited research also into description. Some more related work on Bibliometric has been provided in [7-11].

Proposed Work: To assess the impact of someone research, we need three parameters; (1) total number of citation per article, (2) year of publication, (3) co-authors. We assume that we have the statistics of the total number of citations of an article. We have collected this data through numerous databases like ISI, Researcher ID, Google Scholar, Scopus and web of science etc. We are also aware of the year of publication of those articles and the total number of co-authors per paper. Through our method the possible outcomes are; complete number of papers published by a particular scholar, total number of citation received by an author of those articles, statistical measurements (Mode, mean and median) per paper.

GINI coefficient: We have evaluated Gini's coefficient of each paper. The use of Gini's coefficient to assess someone research work has been described in [12]. Basically, the Gini's coefficient is used to compute statistical dispersal planned to symbolize the salary of a country's citizens, and is the utmost universally used calculator of disparity. It is a single number which is used to measure the grade of dissimilarity in a distribution. Normally it is described mathematically conferring to the Lorenz curve that is used to plot the fraction of the entire revenue of the populace (y axis) that is collectively received by the bottommost x of the populace as shown in figure 1[13]. The line at 45 degrees therefore signifies absolute equivalence of incomes. It is a proportion of the area that stays among the line of equality and the Lorenz curve (shown by blue color arc in the figure 1) above the complete area below the line of equality (shown by green color line in the figure 1).

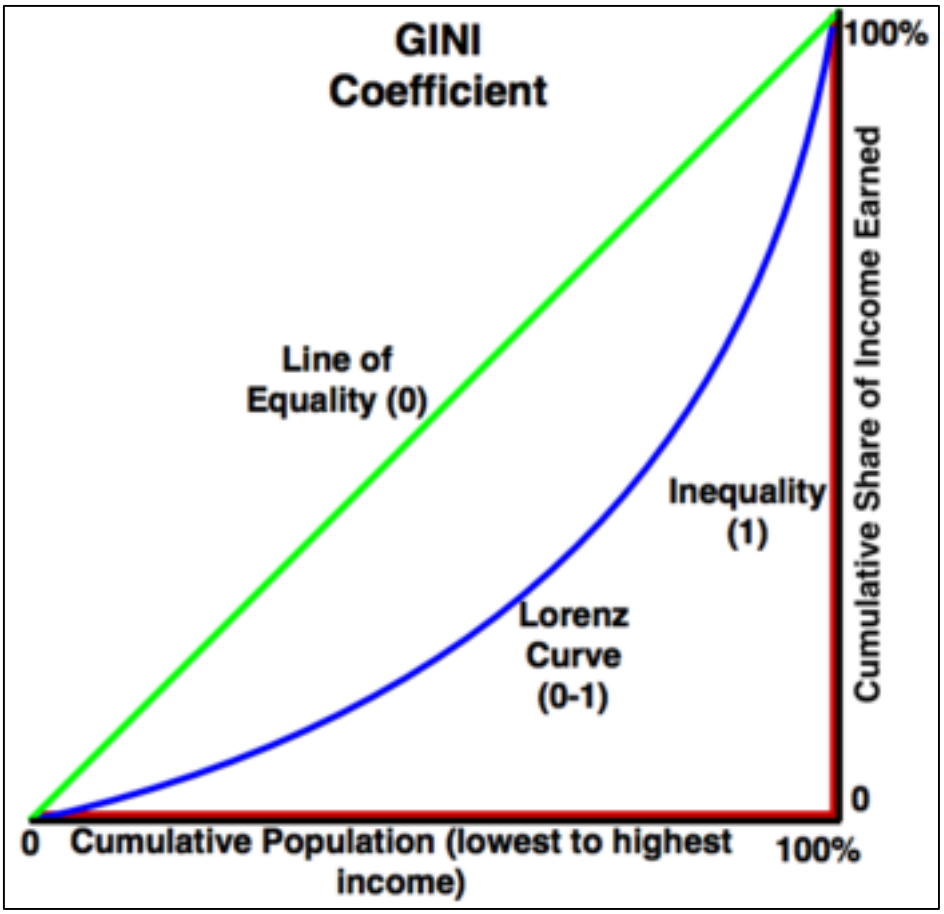

Figure 1: GINI coefficient for poulation

Bibliometric indicators: The following indicators have been implemented and analyzed for numerous statistics.

(i) **Citations indicators**: there are many indicators of citation analysis, few of them are as follows:

(a) **h-index and related parameters; a, m and delta-h:** The h-index was suggested by J.E. Hirsch[14]. It is described by: let a scholar has index *h* iff h of his N articles have minimum h citations respectively, and the remaining (N-h) articles have less than h citations respectively. Its objective is to assess the growing effect of a scholar's outcome through monitoring the number of citation his research has obtained. Publish computes and shows the proper h index, its related rate constant a, and the rate parameter m.

(b) **g-index, delta-g**: The g-index is used for enumerating efficiency in science according to publication record. The g-index was suggested by Leo Egghe [15]. It is described by: provided a collection of papers graded in declining order of the total of citations that they obtained. The g-index is the single biggest number which is maintained by the highest g papers obtained simultaneously. Its objective is to increase the h-index by providing more influence to highly-cited editorials.

(c) **A-index:** The A-index is the Average number of citations of the articles in the h-score [16].

(d) **h2-index**: A scholar's h2 index has been described as the maximum number such that her h(2) most-cited articles obtained for at least [h(2)]2 citations [17].

(e) **e-index**: [18] It is the square root of the redundant of citations in the h-set past h2, means, past the theoretic smallest needed to achieve a h-index of 'h'. The objective of the e-index is to distinguish among scholars with analogous h-indices but dissimilar citation designs.

(f) **Normalized h-Index**: This Modern h-index was anticipated by Antonis et al.[19]. It provides an age based weightage to every cited article (automatically). It is based on the parameterization lesser weight to previous papers. E.g. let gamma=4 and delta=1. It denotes that for a paper distributed in the present year, its citations lead to four times. For a paper distributed four years before, its citations lead to simply one count. For a paper available six years before, its citations credited to 4/6 times, and so on.

(ii) Years weighted indicators:

(a) **Age weighted citation rate (AWCR) and AR-index:** The age-weighted citation rate was established by [20]. It determines the number of citations to a complete form of the research accustomed for the period of every separate article. It is an period based citation rate in which the total count of citations for a specific article is divided by the period of that article. Next, the AR-index is described as the square root of the total of whole period based citation sums above whole articles which are added to the h-index.

(iii) Authors weighted indicators

(a) **Individual h-index (hI-index):** It was designed by Batista [21]. It divides the standard h-index by the average number of scholars of the papers which influence the h-index. As it lessen the consequences of co-authors, it is known as hI.

(b) **Individual h-index (hI, norm-index):** Harzing has implemented a substitute for individual h-index; however it uses a changed method. Initially it normalizes the total number of citations for every article by dividing the total citations by the total authors of that article and next compute the hI. Here norm index means that the normalized citation counts.

(c) **Multi-authored h-index (hm-index):** It was proposed by Michael Schreiber [22]. This technique usages miniscule paper counts rather than lessened citation counts to measure the mutual writer ships of articles and then decides the multi-authored hm index rendering to the subsequent operative grade of the articles utilizing the straight citation counts.

(iv) **Years and Writers weighted indicators:** The per-writer age-weighted citation factor is identical to the simple WCR, however is normalized to the total writers for every article.

Experimental Results: We have made several experiments. Below are the results for the data with 95 % confidence interval. Let the citation received by some research articles by an author/ Writer and co-authors are as follows; 12 8 4 0 5 7 3 2. Let the respective years of publication are as follows; {2012 2013 2014 2015 2016 2017 2018 2019}. Let the number of co-authors for these articles are: {6 4 5 3 8 7 6 5} respectively. The below results show the detailed statistics of Bibliometric and citations.

Complete number of articles: 8

Complete figure of citations: 41

Min Citation Count: 0

Max Citation Count: 12

Mode: 0

Median: 4.5

95% confidence interval: 3.0 - 8.0

Mean: 5.1

Variation coefficient (CV): 74.07%

Adjusted Variation coefficient (CV'): 76.39%

Gini's coefficient: 0.39

Years: 7

Years of printing first: 2012 last: 2019

Articles/year Min: 1 Max: 1

Mode of articles per year: 1

Median of articles per year: 1

95% confidence interval: 1.0 - 1.0

Mean: 1.0

SEM: 0.00

Writers Min: 3 - Max: 8

Mode: 5

Median: 5.5

95% confidence interval: 7.0 - 4.0

Mean: 5.5

SEM: 0.57

95% confidence interval: 4.4 - 6.6

Citations per Writer: 7.3

Bibliometric indicators

Citations indicators

The h-index: 4 a: 2.56 m: 0.57 Delta-h: 1

The g-index: 6 Delta-g: 8

The A-index: 8.00

the h2-index: 2

The e-index: 4.0

the normalized h-index: 0.50

Years weighted indices

The Contemporary h-index (hc-index): 5 a: 1.66

the Age-weighted citation rate (AWCR): 6.23

The AR-index: 2.50

The Age-weighted citation rate (AWCR): 10.39

Harzing's AR-index: 3.22

Writers weighted indices

The hI-index: 0.64

The hI, norm-index: 2.00

The hm-index: 7.00

The figure 2 shows the GINI coefficient for distribution of citations over a period 2012 to 2019. Figure 3 shows H2 index for same data. Figure 4 shows G2 index for same data and period. Figure 5 shows Hc index for same data and period. Figure 6 shows HI norm index for same data and period. Figure 7 shows Hm norm index for same data and period. Figure 8 shows h-index for same data and period.

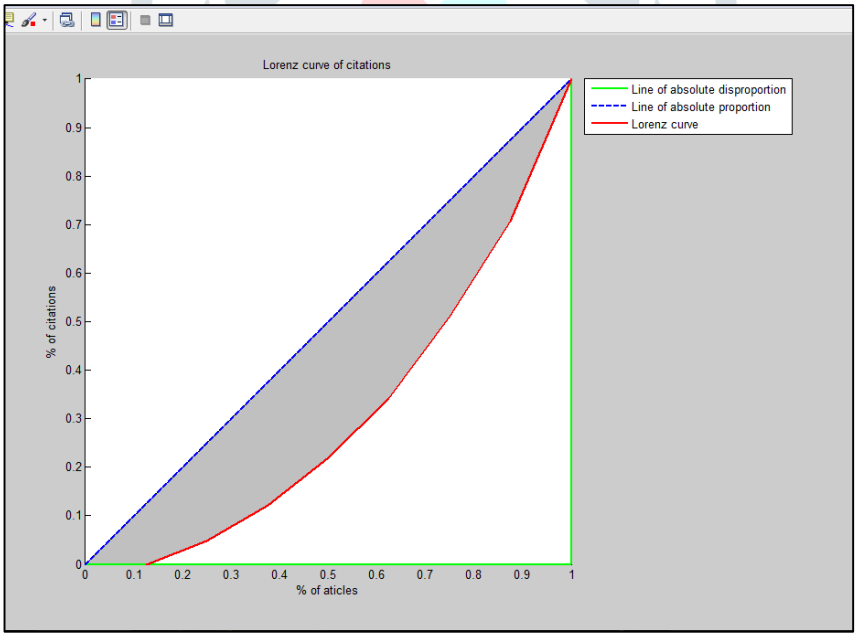

Figure 2: GINI coefficient for distribution of citations over a period

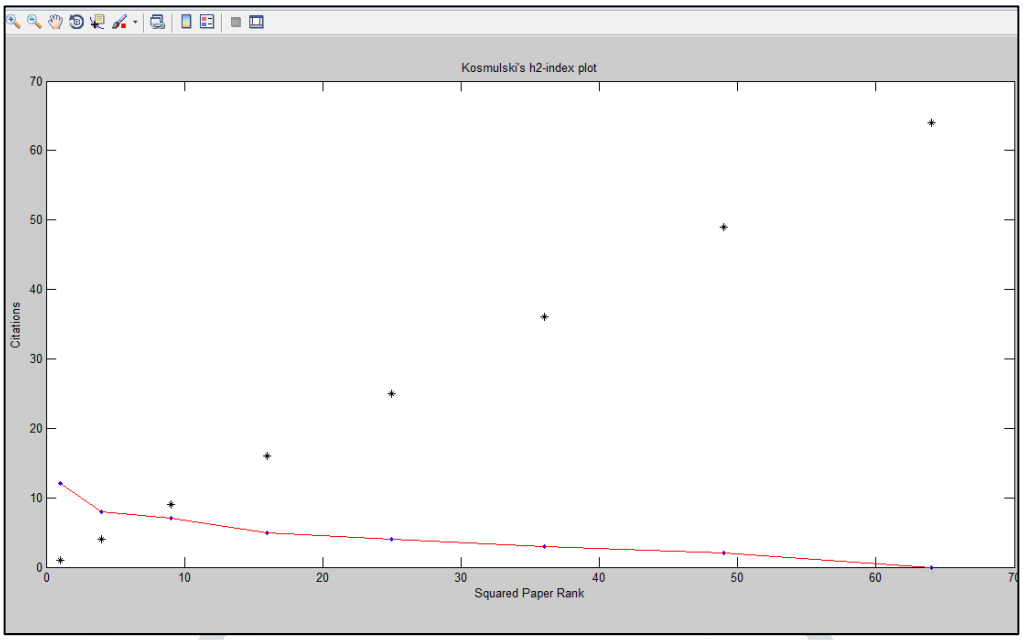

Figure 3: H2 Index

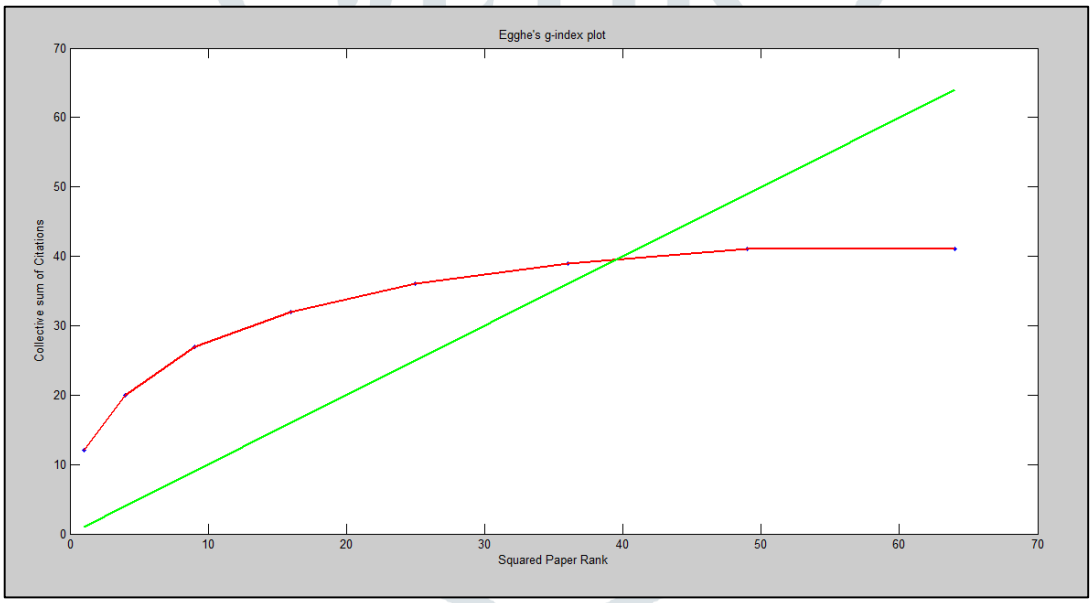

Figure 4: G2 Index

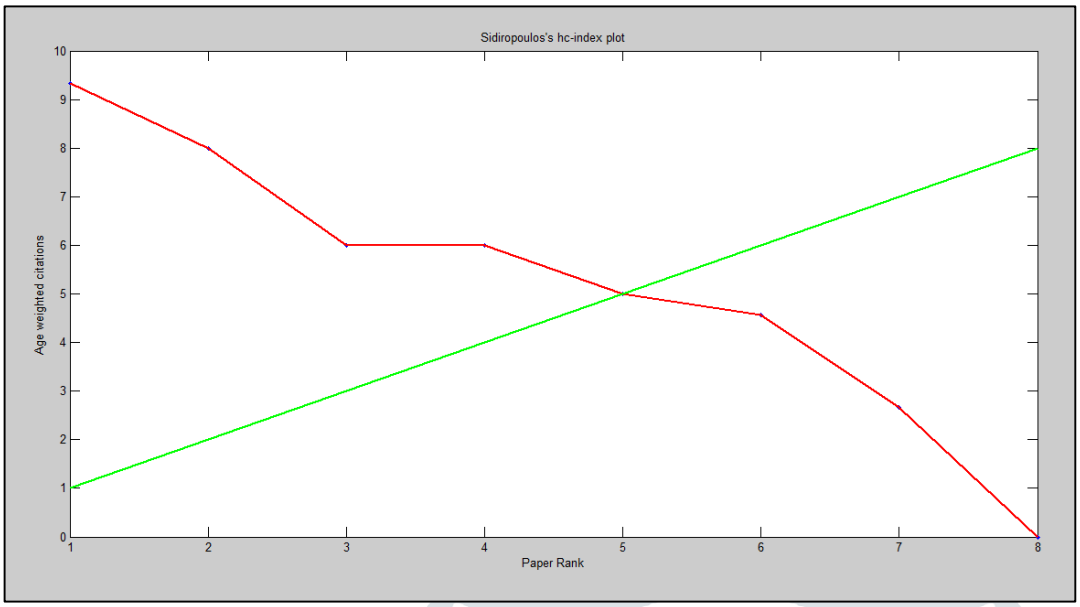

Figure 5: Hc Index

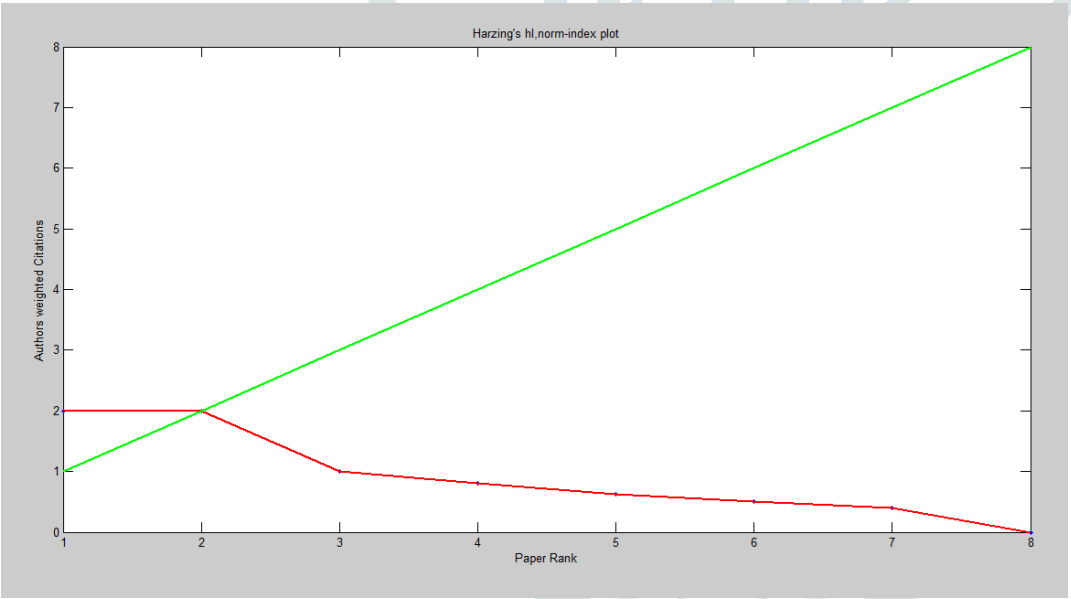

Figure 6: Hi- norm Index

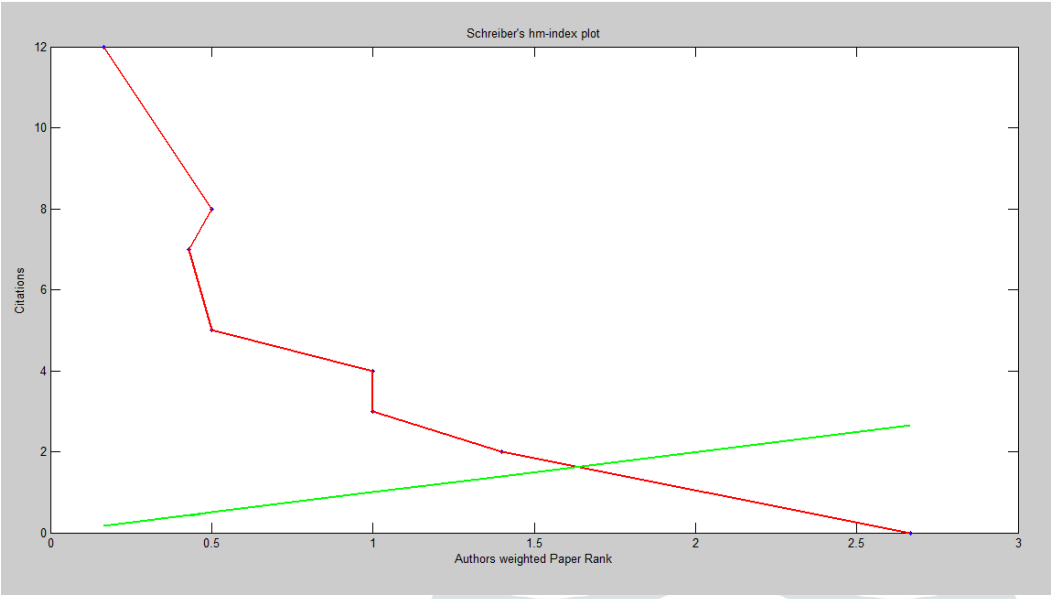

Figure 7:hm-Index

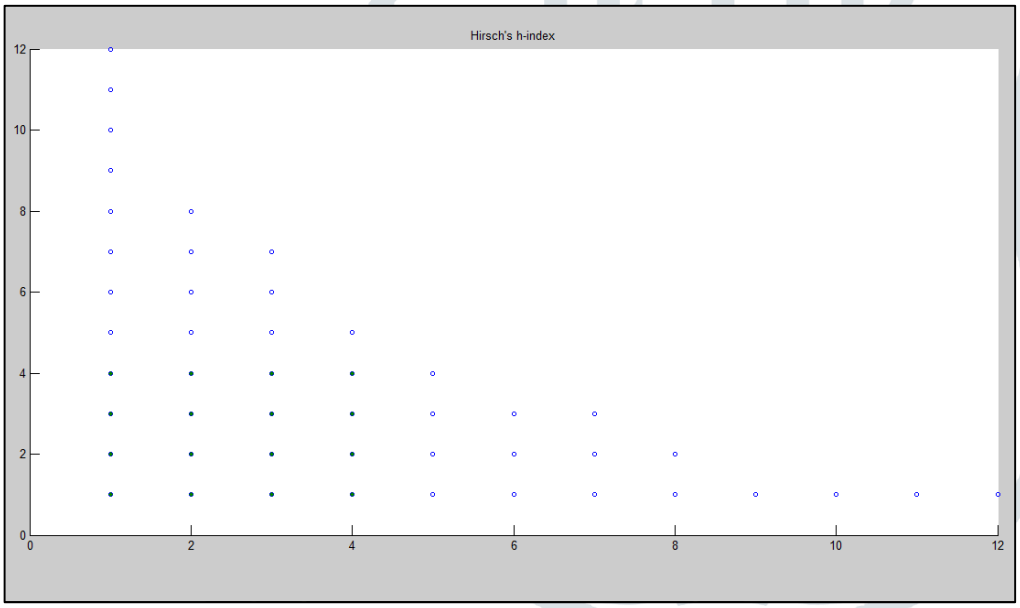

Figure 8: H Index

CONCLUSIONS

Citation indexing has been describes as a main example of an idea prior to its time. Since the formation of the Internet, the time of citation has finally arrived. The mechanism and addition of prevailing Bibliometric tools in international publishing atmosphere must finally leads to the growth of a collection of metrics for improved seizing the total number of manners of intellectual effect inside subject specific societies. There are lots of Bibliometric tools available for the measurement of impact of an individual's research in a particular filed. Each of the tool is good in one or another way.

REFERENCES

[1] Blaise Cronin, Bibliometrics and beyond: some thoughts on web-based citation analysis, Journal of Information Science, Volume: 27 issue: 1, page(s): 1-7, February 1, 2001.

[2] Thomas J. Pehlan, Bibliometrics and the evaluation of Australian sociology, Volume: 36 issue: 3, page(s): 345-363, December 1, 2000.

[3] Priya, Kirti Bhati, Rohini Sharma, Shalini Bhadola, An Overview on Prevailing Bibliographic and Citation Analysis Tools, IJSRD - International Journal for Scientific Research & Development| Vol. 7, Issue 03, 2019, pp. 112-116.

[4] Yinian Gu, Global knowledge management research: A bibliometric analysis, Scientometrics, October 2004, Volume 61, Issue 2, pp. 171–190.

[5] Lars Frode Frederiksen, Disciplinary determinants of bibliometric impact in Danish industrial research: Collaboration and visibility, Scientometrics, October 2004, Volume 61, Issue 2, pp. 253–270.

[6] M.H. MacRoberts and B.R. MacRoberts, Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences, JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 61(1):pp.1–13, 2010.

[7] Björn Hammarfelt, Beyond Coverage: Toward a Bibliometrics for the Humanities, Research Assessment in the Humanities pp 115-131, Springer, 2016

[8] "Bibliometrics". OECD Glossary of Statistical Terms. OECD. 28 March 2013. Retrieved 5 January 2015.

[9] De Bellis, Nicola (2009). Bibliometrics and citation analysis: from the Science citation index to cybermetrics. Scarecrow Press. p. 417. ISBN 978-0-8108-6713-0.

[10] Yeung AWK, Heinrich M, Atanasov AG. Ethnopharmacology-A Bibliometric Analysis of a Field of Research Meandering Between Medicine and Food Science? Front Pharmacol. 2018 Mar 15;9:215. doi: 10.3389/fphar.2018.00215.

[11] Alon, Ilan; Anderson, John; Munim, Ziaul Haque; Ho, Alice (2018). "A review of the internationalization of Chinese enterprises". Asia Pacific Journal of Management. 35 (3): 573–605. doi:10.1007/s10490-018-9597-5.

[12] Tsair-Wei Chien, Julie Chi Chow, Yu Chang, Willy Chou, Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names, Systematic Review and Meta-Analysis, 97:39, 2018, pp.1-11.

[13] https://energyeducation.ca/encyclopedia/Gini_coefficient.

[14] An index to quantify an individual's scientific research output, arXiv:physics/0508025 v5 29 Sep 2005.

[15] Theory and practice of the g-index, Scientometrics, Vol. 69, No 1 (2006), pp. 131-152.

[16] Jin, B. H. h-Index: An evaluation indicator proposed by scientist. Science Focus, 1(1), 8–9 , 2006.

[17] M. Kosmulski, A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter, 2(3), 4–6, 2006.

[18] C.T. Zhang, C.T. The e-index, complementing the h-index for excess citations, PLoS ONE, Vol 5, Issue 5, May 2009.

[19] Antonis Sidiropoulos, Dimitrios %Katsaros, and Yannis Manolopoulos, Generalized h-index for disclosing latent facts in citation networks, arXiv:cs.DL/0607066 v1 13, Jul 2006.

[20] Bihui Jin's note The AR-index: complementing the h-index, ISSI Newsletter, 2007, 3(1), p. 6.

[21] Pablo D. Batista , Monica G. Campiteli, Osame Kinouchi, and Alexandre S. Martinez. Is it possible to compare researchers with different scientific interests?, Scientometrics, vol. 68, No. 1 2006, pp. 179-189.

[22] Michael Schreiber, A modification of the h-index: the hm-index accounts for multi-authored manuscripts, J. Informetrics 2, 211-216 (2008), arXiv:0805.2000 [physics.soc-ph.