

Citation Inequality Among Top NIRF-Ranked Universities: A Gini Coefficient Analysis

Simya K* and Rupesh Kumar A

Department of Library and Information Science, Tumkur University, Tumakuru - 572 103, India

**E-mail: simyamohan97@gmail.com*

ABSTRACT

This study investigated the inequality in citation distribution of the publications produced by top ten universities ranked under the National Institutional Ranking Framework (NIRF) 2024. The analysis aimed to understand the extent of evenness in the distribution of citations across academic publications which is crucial for accurately assessing scholarly impact. Using the affiliation organisation field in the advanced document search in Scopus, a search for publication and citation data of each university was carried out. The search results were refined for the period 2021-2023 to align period of publication and citation data with that of NIRF rank data. The resultant dataset of 56791 publications garnering 725925 citations for period 2021-2023 was exported from Scopus database. Data was analysed using Microsoft Excel. Calculations and visualisations were performed using R statistical analysis software. Gini coefficient, an index to measure the degree of inequality, was used to discover the degree of inequality of citations across publications. The overall Gini coefficient value of 0.6458 revealed a high degree of citation inequality among universities, indicating the concentration of higher number of citations within a small number of publications. Gini coefficient values of open access and non-open access publications were 0.6683 and 0.6152 respectively, highlighting a higher degree of citation inequality among open access publications. No significant association was found between research and publication practices score and Gini coefficient values. In-depth understanding of citation inequality can provide deeper insights on the characteristics of citations in terms of evenness of their distribution and can help uncover a phenomenon where a small number of publications bear the 'burden' of enhancing the citation impact of the entire institution.

Keywords: Citation inequality; Gini coefficient; NIRF universities; Scholarly productivity

NOMENCLATURE

G	: Gini coefficient
N	: Total Number of Publications
X_k	: Cumulative proportion of publications up to k^{th} ranked publication
Y_k	: Cumulative proportion of citations received by k^{th} publication

1. INTRODUCTION

Citations are the building block of academic world which signify the relevance of existing research or invention and are treated as a reflection of how knowledge compiles, integrates and conveys to stimulate new ideas and discoveries¹. Citation analysis has been the core of evaluating research productivity and impact. Thus, citations serve as the foundation for assessing the impact of an individual author/institution/journal. The essence of citations can be used to link research to its intellectual heritage providing insight in to its historical framework and tracing the impact of academic research contributions².

Development of citation databases such as Scopus, Web of Science, Google scholar; availability of full text resources and modern technological advancements have strengthened the growth of citation analysis in terms of citation classifications, citation sentiment analysis, citation summarisation, and citation-based recommendation³.

Researchers are concerned about the limitations of citations in measuring the research impact as it focuses on quantity rather than quality. Citations also have the ability to endorse scientific monopoly in academia, creating a slowdown in true research progress². These factors motivated researchers to unravel the hidden potential of citations to magnify the research impact assessment system. Citation intent analysis aims evaluate the authors purpose behind citing a paper⁴. Citation sentiment analysis has been getting increasing attention among researchers as it analyses authors' sentiment within scientific citations⁵. Classifying citations based on polarity (positive, negative and neutral) can help funding agencies and committees at institutional level to evaluate research performance with greater precision⁶. Asymmetrical distribution of citation among set of publications (or, institutions or authors) is referred to as Citation inequality, which is

a growing concern over the existing research evaluation system⁷. Uneven distribution of citations diminishes the visibility of quality research, so it is important to have an unbiased citation system which provides equal recognition for all scholarly contributions.

1.1 Citation Inequality

Citation analysis has assumed growing significance over the years. But the uneven distribution of citations across publications has become a critical matter of effect on scholarly impact assessment. Thus, citation inequality is a phenomenon of small number of papers or authors receiving a disproportionately large share of citations, while the majority of papers or authors receive little or none⁸. Garfield⁹ highlighted the skewness in citations stating that only 20 % of papers attained over 80 % citations while others are scarcely cited. This unequal distribution of citations influences average citation count which might distort the overall citation pattern¹⁰.

1.1.1 Gini Index

The Gini index was developed by Italian statistician Corrado Gini in 1912. It was used as a significant parameter to measure the socioeconomic inequality. Later, its application was integrated to diverse disciplines including informetrics, especially for citation analysis¹¹. Researches were carried out to measure the citation inequality at geographical level,¹² author level¹³ and institutional level¹⁴. Gini Coefficient for citation distribution can be calculated using Eqn (1): (Where for each article k , cumulative proportion of articles is X_k and cumulative proportion of citations is Y_k)

$$G = 1 - \sum_{k=1}^N (Y_k - Y_{k-1})(X_k + X_{k-1})$$

Eqn (1). Mathematical formula for calculating Gini Coefficient

Table 1. Degrees of inequality based on gini coefficient value

Gini coefficient value	Degree of inequality
0	Perfect equality / No inequality
0.1 to 0.3	Low degree of inequality
0.3 to 0.6	Medium degree of inequality
> 0.6	High degree of inequality
1	Complete inequality

Table 1 presents the Gini coefficient value and its respective degree of inequality. Theoretical value of Gini Coefficient ranges between 0 and 1, which represent perfect level of equality and inequality respectively. Thus, Gini coefficient assesses the degree of inequality of citations across publications and can rebuild the traditional research impact assessment system.

1.2 NIRF Ranking

The National Institutional Ranking Framework (NIRF) is an initiative by Ministry of Education, Government of India to rank higher education institutions in the

country, launched on 29th September 2015¹⁵. Research and Professional Practices, one of the parameters used to assess the overall performance, evaluates the quality, quantity and impact of research conducted by institutions. The scholarly impact of institutions is measured through traditional citation-based data. The effect of uneven distribution of citations appearing in institutional publications should be precisely assessed to revamp the research evaluation methods.

1.3 Present Study

Universities stand as the pillars of formulating knowledge in contemporary societies¹⁴. Institutional research productivity evaluation has attracted the attention of researchers, particularly those engaged in bibliometric studies¹⁶⁻¹⁹. In this scenario, it is well known that some universities are more productive than others in terms of the average number of citations that these publications receive. The citation inequality opposes the universality in citation distributions, as higher number of citations are acquired by only a few papers while rest of them are rarely cited or not cited at all. Inequality in citation distribution will have a great impact on productivity of academic institutions. While considering ranking of institutions citation inequality can make a big difference on scholarly productivity. Hence, the present study investigated the skewness in distribution of citations across top ten NIRF-ranked universities to identify the extent of citation inequality among their publications.

2. OBJECTIVES

- To understand the publication and citation patterns of the top ten NIRF-ranked universities.
- To evaluate the degree of inequality in the distribution of citations among these institutions.
- To identify the difference in degree of inequality in the distribution of citations across open access and non-open access publications.

3. METHODOLOGY

The data for the present study was collected from two sources, namely, NIRF website and Scopus database. Data on NIRF rank, total score and Research and Publication Practices (RP) score were sourced from NIRF website. Top ten NIRF-ranked universities for the year 2024 were listed out with their score. NIRF 2024 used the past three years' data (2021-2023) to assess the institutions for their excellence.

The Scopus database was used for the present study for collecting the data on publications and citations. The reliability of Scopus has made it a significant bibliometric data source for extensive analyses in research assessments, tracing research landscape, science policy evaluations, and university rankings²⁰. Advanced document search was performed in Scopus using Affiliation Organisation (AFFILORG) field. The exact name of each university was entered in this field. The search results were refined

for the period from 2021 to 2023. All document types available in Scopus were considered. Documents with zero citations were excluded from the study.

Accordingly, a dataset of 56791 publications for the period 2021-2023 was exported and analysed using Microsoft Excel. R statistical analysis software²¹ was used to calculate the Gini coefficient to measure the degree of inequality in the distribution of citations among universities and to plot Lorenz curve to graphically represent the same.

4. RESULTS AND DISCUSSION

4.1 Scholarly Productivity of Top NIRF Universities

NIRF uses five parameters: Teaching, Learning and Resources, Research and Professional Practices, Graduation Outcomes, Outreach and Inclusivity, and Perception to rank institutions across the country¹⁵. Table 2 shows the top ten NIRF universities with their total score and Research and Professional Practices (RP) score sourced from NIRF website¹⁵. RP score is a composite score calculated by considering five parameters, namely, Combined metric for publications, Combined metric for quality of publications, IPR and Patents: Published and granted, Footprint of Projects and Professional practice, and Publications & Citations in SDGs. The number of publications and citations were considered for the period from 2021 to 2023. All the Universities together contributed a total of 56791 publications and accumulated 725925 citations over the period of 2021 to 2023, averaging 12.78 citations per publication. Vellore Institute of Technology, the tenth ranked university had the highest number of publications (9040) and citations (120816), while Amrita Vishwa Vidyapeetham, seventh ranked university had the lowest number of publications (2178) and citations (18260). In terms of average citations per publication,

Jamia Milia Islamia had the highest (16.05) and Amrita Vishwa Vidyapeetham had the lowest (8.38). Higher RP score attained by the Indian Institute of Science (86.5) and Vellore Institute of Technology (61.95) is substantiated by the greater number of publications and citations received by them during these years. Citation analysis highlights the impact of research performance of Vellore Institute of Technology, Manipal Academy of Higher Education and Banaras Hindu University in disseminating insightful research among scholarly community.

4.2 High Degree of Citation Inequality

Table 3 presents the calculated Gini coefficient value for top ten NIRF ranked Universities. Gini coefficient value for universities ranged between 0.5995 and 0.6926. Manipal Academy of Higher Education had the highest Gini coefficient value (0.6956) followed by Banaras Hindu University (0.6926) while Amrita Vishwa Vidyapeetham had the lowest Gini coefficient value of 0.5995 which represented a medium degree of inequality. The remaining nine universities had a Gini coefficient value greater than 0.6 representing a high degree of inequality, indicating that greater number of citations were concentrated in small number of publications.

The overall Gini coefficient value of 0.6458 revealed a high degree of inequality in the distribution of citations among publications of top ten NIRF ranked Universities, exhibiting an uneven distribution of citations among institutional publications.

4.3 Lorenz Curve for Citation Distribution

Figure 1 demonstrates the Lorenz curve for overall citation distribution of top ten NIRF ranked universities. The Lorenz Curve was devised by the American statistician Max O. Lorenz in 1905²². Present study used Lorenz

Table 2. Scholarly productivity of NIRF ranked universities

NIRF rank	University	Total score*	RP score*	No. of publications	Total citations	ACPP**
1	Indian Institute of Science, Bengaluru (IISc)	83.29	86.5	7782	90163	11.59
2	Jawaharlal Nehru University (JNU)	69.8	45.27	3372	41231	12.23
3	Jamia Millia Islamia (JMI)	68.11	49.28	3572	57319	16.05
4	Manipal Academy of Higher Education, Manipal (MAHE)	67.18	55.94	7120	95788	13.45
5	Banaras Hindu University (BHU)	66.05	51.01	6026	90641	15.04
6	University of Delhi (UoD)	65.9	56.42	7513	87285	11.62
7	Amrita Vishwa Vidyapeetham, Coimbatore (AVV)	65.73	47.82	2178	18260	8.38
8	Aligarh Muslim University (AMU)	65.57	45.91	5218	68326	13.09
9	Jadavpur University (JU)	65.39	51.35	4970	56096	11.29
10	Vellore Institute of Technology (VIT)	64.79	61.95	9040	120816	13.36
Total				56791	725925	12.78

*Source: NIRF Website; **ACPP: Average Citations Per Publication

curve to graphically represent the inequality in citation distribution. The line between the origin of the coordinates and the corresponding vertex is the line of perfect equality. The area between Lorenz Curve and the line of equality shows the exact degree of inequality. Thus, the less deviation from the line of perfect equality represents a more even distribution²².

4.4 Citation Inequality Across Open Access Publications

As indicated in table 4, all the ten universities together produced a total of 21140 open access publications and attained a total of 357120 citations. Average number of open access publications per institution was 2114. Manipal Academy of Higher Education had the highest number of open access publications (4075) and citations (66245). Banaras Hindu University stood second in citations (50379), followed by Indian Institute of Science (47421). Mean value of citations received by these institutions for open access publications was 16.89×8 .

Overall Gini coefficient value for open access publications was 0.6683 which represents a high degree of inequality in the distribution of citations for open access publications among top ten NIRF ranked Universities. Banaras Hindu University had the highest Gini coefficient

Table 3. Gini coefficient values of top ten NIRF ranked universities

University	Gini coefficient value
Indian Institute of Science, Bengaluru	0.6153
Jawaharlal Nehru University	0.6471
Jamia Millia Islamia	0.6476
Manipal Academy of Higher Education, Manipal	0.6956
Banaras Hindu University	0.6926
University of Delhi	0.6375
Amrita Vishwa Vidyapeetham, Coimbatore	0.5995
Aligarh Muslim University	0.6148
Jadavpur University	0.6103
Vellore Institute of Technology	0.6261
Overall gini coefficient value	0.6458

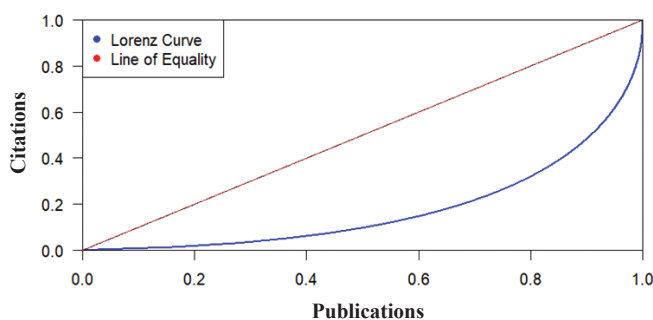


Figure 1. Lorenz curve for overall distribution of citations. The area between line of equality and lorenz curve indicates the degree of inequality in the citation distribution. larger the area, greater the inequality.

value for open access publications (0.7548) followed by Manipal Academy of Higher Education (0.7224). Moreover, all the ten universities had a Gini coefficient value greater than 0.6 indicating a significant degree of uneven distribution of citations for open access publications. The Gini coefficient value for overall distribution of citations for open access publications is illustrated using Lorenz curve in Fig. 2.

4.5 Citation Inequality Across Non-Open Access Publications

Table 5 shows that a total of 35651 non-open access publications were produced by all the ten universities which attained a total of 368805 citations. Vellore Institute of Technology (5976) produced the highest number of non-open access publications followed by University

Table 4. Gini coefficient for open access publications

University	Open access (OA) publications	Citations	Gini coefficient (OA)
Indian Institute of Science, Bengaluru	3333	47421	0.6211
Jawaharlal Nehru University	1105	19567	0.6519
Jamia Millia Islamia	1212	27747	0.6589
Manipal Academy of Higher Education, Manipal	4075	66245	0.7224
Banaras Hindu University	2120	50379	0.7548
University of Delhi	2580	40310	0.6404
Amrita Vishwa Vidyapeetham, Coimbatore	581	6906	0.6071
Aligarh Muslim University	1943	30574	0.608
Jadavpur University	1127	18392	0.6498
Vellore Institute of Technology	3064	49579	0.6163
Overall	21140	357120	0.6683

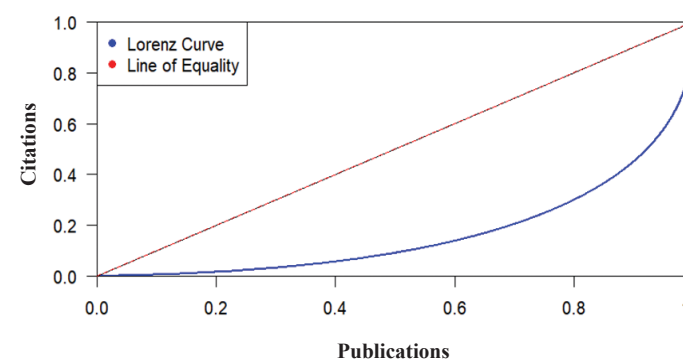


Figure 2. Lorenz curve for overall distribution of citations for open access publications.

of Delhi (4933) and Indian Institute of Science (4449). Vellore Institute of Technology(71237), University of Delhi(46975) and Indian Institute on Science(42742) received higher citations for non-open access publications. Average number of citations received by a non-open access publications is 10.34.

Overall Gini coefficient value for non-open access publications was 0.6152 indicating a high degree of inequality in the distribution of citations across non-open access publications. Manipal Academy of Higher Education had the highest Gini coefficient value (0.6295) for non-open access publications followed by Vellore Institute of Technology (0.6266). Three out of ten universities namely, Indian Institute of Science, Amrita Vishwa Vidyapeetham and Jadavpur University had Gini coefficient value between the range 0.3 and 0.6 illustrating a moderate degree of citation inequality for non-open access publications. The

Table 5. Gini coefficient for non-open access publications

University	Non-open access (NOA) publications	Citations	Gini coefficient (NOA)
Indian Institute of Science, Bengaluru	4449	42742	0.5989
Jawaharlal Nehru University	2267	21664	0.6242
Jamia Millia Islamia	2360	29572	0.6208
Manipal Academy of Higher Education, Manipal	3045	29543	0.6295
Banaras Hindu University	3906	40262	0.6017
University of Delhi	4933	46975	0.6209
Amrita Vishwa Vidyapeetham, Coimbatore	1597	11354	0.5795
Aligarh Muslim University	3275	37752	0.6134
Jadavpur University	3843	37704	0.5847
Vellore Institute of Technology	5976	71237	0.6266
Overall	35651	368805	0.6152

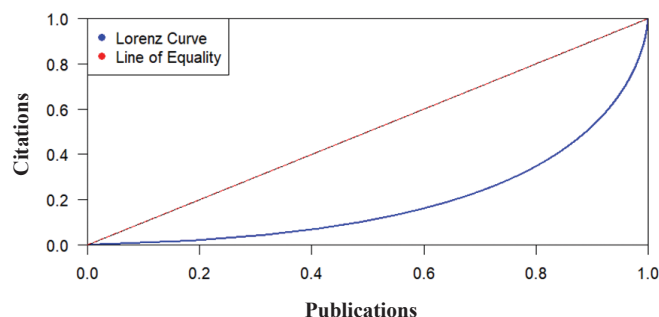


Figure 3. Lorenz curve for overall distribution of citations for non-open access publications.

Gini coefficient value for overall distribution of citations for non-open access publications is demonstrated using Lorenz curve in fig. 3.

5. CONCLUSION

This study explored a phenomenon where a small number of publications bear the ‘burden’ of enhancing the citation impact of the entire institution. This “citation inequality” among publications produced by top NIRF ranked universities has been illustrated under the present study. The study examined the degree of inequality in distribution of citations using Gini coefficient. The overall Gini coefficient value of 0.6458 revealed a high degree of inequality in the distribution of citations across university publications. Gini coefficient value for universities ranges from 0.5995 to 0.6926 exhibiting a moderate to high degree of inequality. Analysis of distribution of citations among open access and non-open access publications found out a high degree of citation inequality. Overall Gini coefficient value for both open access and non-open access publications were greater than 0.6 indicating an uneven distribution of citations across publications to a great extent. Overall Gini coefficient value of open access publications (0.6683) was higher than that of non-open access publications (0.6152) showing that open access publications exhibit a marginally higher degree of citation inequality. No significant association was observed between RP Score and Gini Coefficient values. A more comprehensive study with a larger sample size may be necessary to further examine this relationship. The study reviewed the effect of inequality measures in the distribution of citations across academic publications which can make a big difference in scholarly impact. We suggest that ranking systems such as NIRF may integrate citation inequality measures in assessing the institutional research impact.

REFERENCES

- Comins JA, Hussey TW. Compressing multiple scales of impact detection by reference publication year spectroscopy. *J Informetr.* 2015 Jul;9(3):449-54. doi: 10.1016/j.joi.2015.03.003
- Kostoff RN. The use and misuse of citation analysis in research evaluation. *Scientometrics.* 1998 Sep;43(1):27-43. doi: 10.1007/BF02458392
- Iqbal S, Hassan SU, Aljohani NR, Alelyani S, Nawaz R, Bornmann L. A decade of in-text citation analysis based on natural language processing and machine learning techniques: An overview of empirical studies. *Scientometrics.* 2021 Aug;126(8):6551-99. doi: 10.1007/s11192-021-04055-1
- Roman M, Shahid A, Khan S, Koubaa A, Yu L. Citation intent classification using word embedding. *IEEE Access.* 2021;9:9982-95. doi: 10.1109/ACCESS.2021.3050547
- Yousif A, Niu Z, Tarus JK, Ahmad A. A survey on sentiment analysis of scientific citations. *Artif Intell Rev.* 2019 Oct;52(3):1805-38. doi: 10.1007/s10462-017-9597-8

6. Abu-Jbara A, Ezra J, Radev D. Purpose and polarity of citation: Towards NLP-based bibliometrics. In: Vanderwende L, Daumé III H, Kirchhoff K, editors. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics. 2013 cited 2025 Feb 22:p. 596–606. Available from: <https://aclanthology.org/N13-1067/>
7. Banerjee suchismita, Ghosh A, Basu B. Exploring citation diversity in scholarly literature: An entropy-based approach. *arXiv*;2024 [cited 2025 Feb 20]. Available from: <http://arxiv.org/abs/2409.02592>
8. Tahamtan I, Safipour Afshar A, Ahamdzadeh K. Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*. 2016 Jun;107(3):1195–225. doi: 10.1007/s11192-016-1889-2
9. Garfield E. The History and Meaning of the Journal Impact Factor. *JAMA*. 2006 Jan 4;295(1):90. doi: 10.1001/jama.295.1.90
10. Kiesslich T, Beyreis M, Zimmermann G, Traweger A. Citation inequality and the journal impact factor: Median, mean, (does it) matter? *Scientometrics*. 2021 Feb;126(2):1249–69. doi: 10.1007/s11192-020-03812-y
11. Sitthiyot T, Holasut K. A simple method for measuring inequality. *Palgrave Commun*. 2020 Jun 4;6(1):112. doi: 10.1057/s41599-020-0484-6
12. Ettarh R. Analysis of citation inequality in Finland and Nigeria using the Lorenz curve. *South Afr J Sci*. 2021 Sep 29 [cited 2025 Feb 22];117(9/10). Available from: <https://sajs.co.za/article/view/10726>
13. Chien TW, Chow JC, Chang Y, Chou W. Applying gini coefficient to evaluate the author research domains associated with the ordering of author names: A bibliometric study. *Medicine (Baltimore)*. 2018 Sep;97(39):e12418. doi: 10.1097/MD.00000000000012418
14. Perianes-Rodriguez A, Ruiz-Castillo J. University citation distributions. *J Assoc Inf Sci Technol*. 2016;67(11):2790–804. doi: 10.1002/asi.23619
15. National institutional ranking framework (NIRF). 2024 [cited 2024 Dec 14]. Available from: <https://www.nirfindia.org/Home/About>
16. Bala A, Kumari S. Research performance of national institutes of technology (NITS) of India during 2001-2010: A bibliometric analysis. *SRELS J Inf Manag*. 2013;50(5):555–72.
17. Kumbar M, Gupta BM, Dhawan SK. Growth and impact of research output of university of Mysore, 1996-2006: A case study. *J Inf Manag*. 2008 Dec 30;39(4):241–57. doi: 10.1633/JIM.2008.39.4.241
18. Satpathy D, Sa M. Research outputs of state government universities of Odisha: A bibliometric study. *Libr Philos Pract*. 2015 Jan 1. Available from: <https://digitalcommons.unl.edu/libphilprac/1309>
19. Wadhwa NK, Tewari DK, Walke R, Yadav AK, Dhawan SM. Bibliometric analysis of NPL papers published during 1981 –1985 and 2001 – 2005: Case study. In: *Proceedings of WIS 2008*.
20. Baas J, Schotten M, Plume A, Côté G, Karimi R. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quant Sci Stud*. 2020 Feb;1(1):377–86. doi: 10.1162/qss_a_00019
21. R Core Team. R: A language and environment for statistical computing (Version 4.4.2), 2024. Available from: <https://www.R-project.org>.
22. Jin J, Wang J, Ma X, Wang Y, Li R. Equality of medical health resource allocation in China based on the gini coefficient method. *Iran J Public Health*. 2015 Apr;44(4):445–57. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441957/>

CONTRIBUTORS

Ms. Simya K is a UGC Junior Research Fellow in the Department of Studies and Research in Library and Information Science, Tumkur University, Tumakuru. Her areas of interest include: Institutional productivity, Research evaluation and Bibliometrics. Her contributions to the present study include: developing the conceptual framework, data collection, analysis and drafting of the manuscript.

Dr. Rupesh Kumar A is an Associate Professor in the Department of Studies and Research in Library and Information Science, Tumkur University, Tumakuru, Karnataka. His areas of interest include: Citation context analysis, Research evaluation and Webometrics. His contributions to the present study include: Ideation, formulation of research design, inferencing and finalisation of the manuscript.