DESIDOC Journal of Library & Information Technology, Vol. 39, No. 2, March 2019, pp. 60-66, DOI : 10.14429/djlit.39.2.14163

Finding Facets of Academic Integrity and Plagiarism through the Prism of a Citation Database

Swapan Kumar Patra[#,*] and Anup Kumar Das[$]

[#]*Tshwane University of Technology, Pretoria, South Africa*
[$]*Centre for Studies in Science Policy, Jawaharlal Nehru University, India*
[*]E-mail: skpatra@gmail.com

ABSTRACT

With the recent growth of university education and the increasing amount of literature available on the web, there is concern among the academicians and others interested about the research ethics and academic integrity. This paper is an attempt to map the global publication and research trends in the academic integrity and plagiarism issues. The paper uses different bibliometrics, scientometrics and social network tools to map the literature growth pattern, subject areas, document types, source types of scholarly literature. It also analyses subject through keywords analysis and identifies core journals, productive institutes and productive countries. The study is trying to situate India's position and its research trends in the global landscape. It is observed that the India's scholarly literature in these areas are increasing in recent years. Although it is not at par with the global literature growth. The keyword analysis observed that Indian literature is more towards computer and ICT related subjects rather than the global research in the medical subject (human is the second most occurring keyword). The study recommends more focused and integrated research approach from Indian researchers.

Keywords: Plagiarism; Academic integrity; Research integrity; Publication ethics; Keyword analysis; Network analysis; India; UGC

1. INTRODUCTION

In recent years, there is a growing interest among the scholars, researchers, general audience and also among other interested parties about academic integrity and plagiarism. This growing interest is because of the expansion of higher education globally and also the result of the standards of professional conduct[1]. The professional integrity of scientists is important to society as a whole and particularly to the disciplines such as different branches of biological science because the subject is very dynamic in nature. These types of high technology subject areas depend heavily on research and development (R&D) and the subject progress very rapidly. The study found that globally there are quite strong and similar norms of professional conduct, but the unethical acts differed in its approaches[2].

Fraudulence in research and publication, be intentional or not, goes much beyond an act of delinquency. The consequences can be dire and far-reaching. Regardless of economic and academic progression, plagiarism remains to be a global concern. The publishing houses, editors and authors are bound by the publication ethics, yet effective retributive tools remain scattered and unable to curtail the epidemic[3].

The university libraries in India and other countries are actively involved in awareness raising and sensitising the university students, researchers, and faculty members how to uphold the good scientific practices and avoid plagiarism.

1.1 Seven Deadly Sins of Plagiarism

In a recent online tutorial titled "Guide on plagiarism and how to avoid it", the KAUST Library in Saudi Arabia highlights "Seven deadly sins of plagiarism" vis-à-vis "Seven academic integrity commandments". "Those are summarily:

- Failure to give proper credit
- Copying materials from Internet without citing it
- Failure to cite even a few words of borrowed language
- Failure to cite an exact quote
- Failure to cite paraphrased ideas
- Failure to provide an accurate citation, and
- Thinking you can get away with plagiarism"[4].

India's efforts were widely discussed in print as well as online media[5]. The premier publications such as *Science* and *Nature* published perspective papers on new regulations in India that have flagged several issues such as non-availability of awareness raising and sensitisation efforts at the local level, inappropriate infrastructure in universities, and the missing curricula that could have empowered the HEI students with original writing skills.

Recently, the Government of India has adopted a new policy and consequently has passed regulation to deal with the issues related to academic integrity and plagiarism[6]. The regulations on "Promotion of academic integrity and

Received : 21 February 2019, Revised : 26 February 2019
Accepted : 28 February 2019, Online published : 11 March 2019

prevention of plagiarism in higher educational institutions" adopted by the University Grant Commission (UGC) has come up with various issues. The policy document further delimits the plagiarism as follows:

- All quoted work reproduced with all necessary permissions and/or attribution
- All references, bibliography, table of content, preface and acknowledgements
- "All generic terms, laws, standard symbols and standard equations" are not considered as plagiarism[7].

In this context, this study is an attempt to map the research areas related to academic integrity and plagiarism. While doing so, the study will map the growth of literature, different forms and outlets of publications, the scholarly journals where the research articles have been published, the productive institutions and so on. This article is also mapping the trend in scholarly publications from India and how it is aligned with the global literature in terms of the number of publications, the journal where it is publishing the research trends (by analysing keywords), the productive institutes and so on.

2. RESEARCH OBJECTIVES

The article is trying to investigate the issues raised above with the following research objectives:

- To trace the growth of literature in this area globally and the growth of literature from India
- To find the subject areas, document types, source types of the literature published both globally and from India
- To map the research trends using keyword analysis tools
- To find the core journals where the maximum number of research articles have been published
- To find the productive institutes globally, as well as from India.

3. METHODOLOGY

For this study, data have been extracted from the Scopus citation database of Elsevier® science. Scopus is the largest abstract and citation database of peer-reviewed scholarly literature. The database collects and index about 22,800 different titles from different sources of more than 5,000 international publishers. Its coverage includes the whole universe of knowledge in a wide variety of subjects (science, technology, medicine, social science, arts, and humanities)[8].

For this study records were searched using the following search string: ((TITLE-ABS-KEY (Plagiari*) OR TITLE-ABS-KEY ("Research Integrity") OR TITLE-ABS-KEY ("Academic Integrity") OR TITLE-ABS-KEY ("Research Misconduct") OR TITLE-ABS-KEY ("Scientific Misconduct") OR TITLE-ABS-KEY ("Publication Ethics") OR TITLE-ABS-KEY ("Academic misconduct")) AND (LIMIT-TO (PUBYEAR,2000-2018)). The search was limited from the publication year 2000 to the publication year 2018.

This search operation retrieved about 11,318 record globally. These records are downloaded from the Scopus databases as 'comma delimited format' and stored in Microsoft Excel sheet for further analysis. The retrieved results are also downloaded in 'Scopus Export Refine' format to check the number of yearly publication records. The Indian publications were extracted from the whole set of global retrieved records and stored in a separate excel sheet to investigate the trends in publications from the Indian scholars. Among the total 11,318 articles, 459 article are published by Indian researchers. The retrieved set of records are used for analysis according to the objective of the study.

The keywords from the articles are separated from the database, and the network map of keywords are plotted using open source social network analysis tools. For drawing the network maps and to get different social network indicators, the following open source network tools, Gephi[9], VOS-viewer[10] and UCINET[11] for windows are used.

4. RESULTS

4.1 Literature Growth

The global literature growth pattern shows that, the scholarly publication in this filed is growing in a linear fashion. From 203 article in 2000, there are about 1,024 article published in 2018. In the Indian case, the number of research publication was about three articles in 2001 and about 66 article in 2018. The maximum number of publications happened in 2016 in both cases. In that year, globally about 1,072 article and 77 article from India has been published. The significant growth of Indian publications has been observed from 2011 onwards (Fig. 1). The literature growth in both cases is highly co-related.

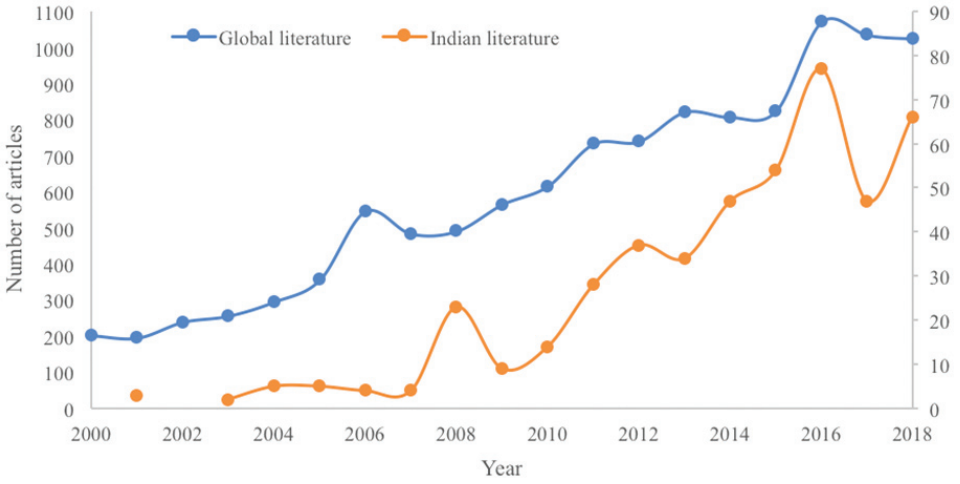

Figure 1. Growth of publications: global vs Indian literature.

4.2 Subject Categories

Scopus® has categorised Subject Areas into 27 major thematic categories according to its own classification. Based on these thematic subject classifications, as seen from the Fig. 2, globally the maximum number of publications are in the areas of medicine (4,125) followed by social sciences (3,317) and computer science (1,927). In the Indian case, the maximum number of publications are from computer science (148) and medicine (148) followed by social sciences (91).

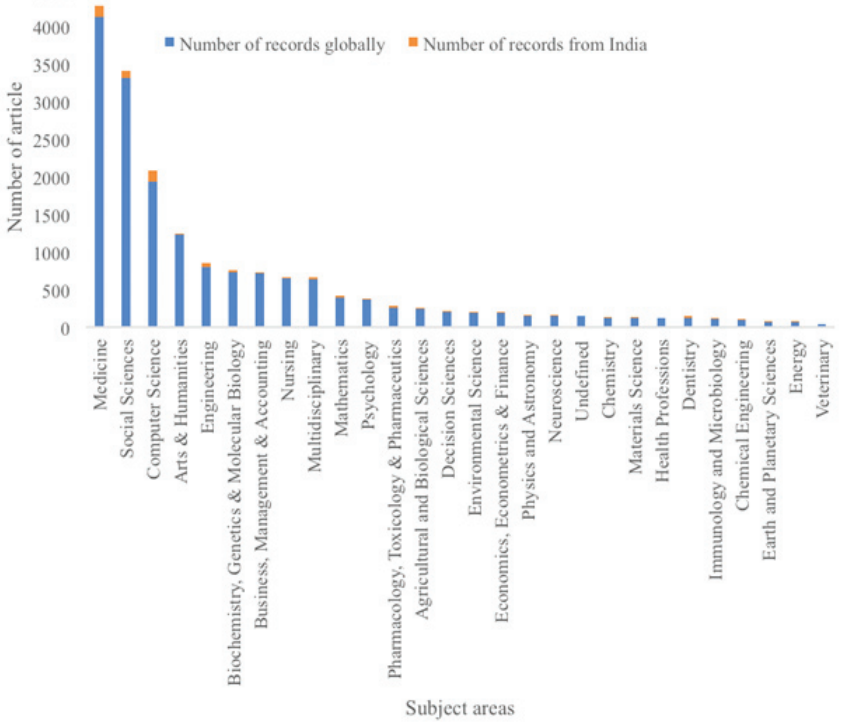

Figure 2. Subject categories as on Scopus database.

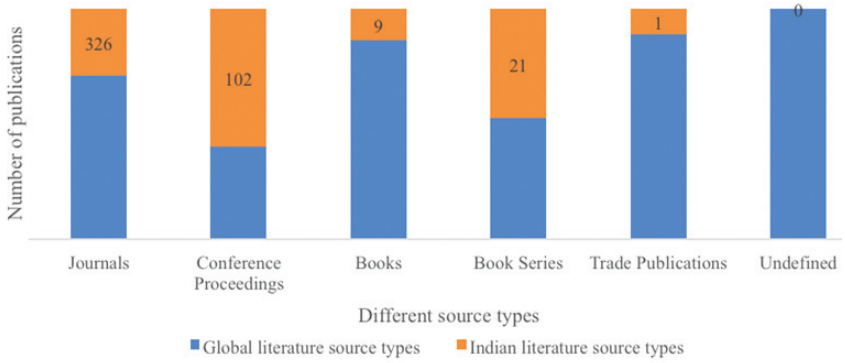

Figure 3. Distribution of literature based on source types.

4.3 Document Types

Scopus in its broad subject coverage included various document types bust mostly from serial publications. Its documents types include articles, books, book chapters, conference proceedings, letters, notes and so on. However, according to Scopus content coverage guide, the database does not cover book reviews and conference meeting abstracts. Hence these types of documents are not in the analysis of this paper.

Distribution of literature based on different document types shows that the maximum number of documents is being published as journal articles. During the 2000-2018 years' period, globally, there are about 4,278 (about 38 percent) articles published as journal articles in different journals followed by 1,560/14 percent) conference papers. From India, there are 132 journal articles (29 percent) and 125 conference papers (27 percent).

4.4 Source Types

Scopus database indexes different serial publications including journals, trade journals, book series and conference materials and so on based on its own content coverage policy. Journal publication is the predominant mode of publication outlet. From the global literature, there are about 9,018 articles published as the journal articles, and from India 102 are the journal articles. However, it is obvious because Scopus covers more journal articles than any other form of publications. Fig. 3 shows the distribution of different source types both from India and globally.

4.5 Keywords Analysis

Since the last century, scientific literature has increased significantly in all sphere of knowledge. Along with the growth of knowledge, scholarly literature from this field has increased significantly. This can be seen from the global literature growth and also the growth of literature from India (Fig. 1). To map the nature of the subject, this paper did keyword analysis. The keyword analysis is done using the index keyword assigned to every article from the retrieved set of data. According to Scopus content coverage guide, index keywords "manually added by the subject experts and index terms for 80 per cent of the titles included in Scopus. These index terms are derived from thesauri that Elsevier owns or licenses and are added in order to improve search recall"[8].

Based on the co-occurrence frequency of pairs of words or phrases, the co-word analysis is used to discover linkages among topics in a research field. This technique, used by several researchers and showed that keyword analysis is a powerful tool for knowledge discovery[12]. Researchers have used co-word analysis as an important method to explore the concept network in different fields[13]. This method shows the relationship between different entities through visual representation maintaining the essential information containing within the data[14]. It is based on the nature of words, which are the important carrier of scientific concepts, idea, and knowledge[15].

The keywords from all articles of all years are collected. The keywords are further standardised to remove some ambiguity and also standardised. A network map of relationships among keywords is drawn and further plotted using social network tools (Fig. 4).

4.6 Centrality Measures of Keywords

Social network analysis (SNA) is a powerful tool originate from mathematical graph theory. This powerful analytical tool is increasingly used in classical sociology and its more recently in almost every sphere of knowledge including the formulation in social, scientific and mathematical research work. Social network analysis has been used to study the relationship among different entities for example, in citation network, authorship, collaboration networks and so on[16]. By

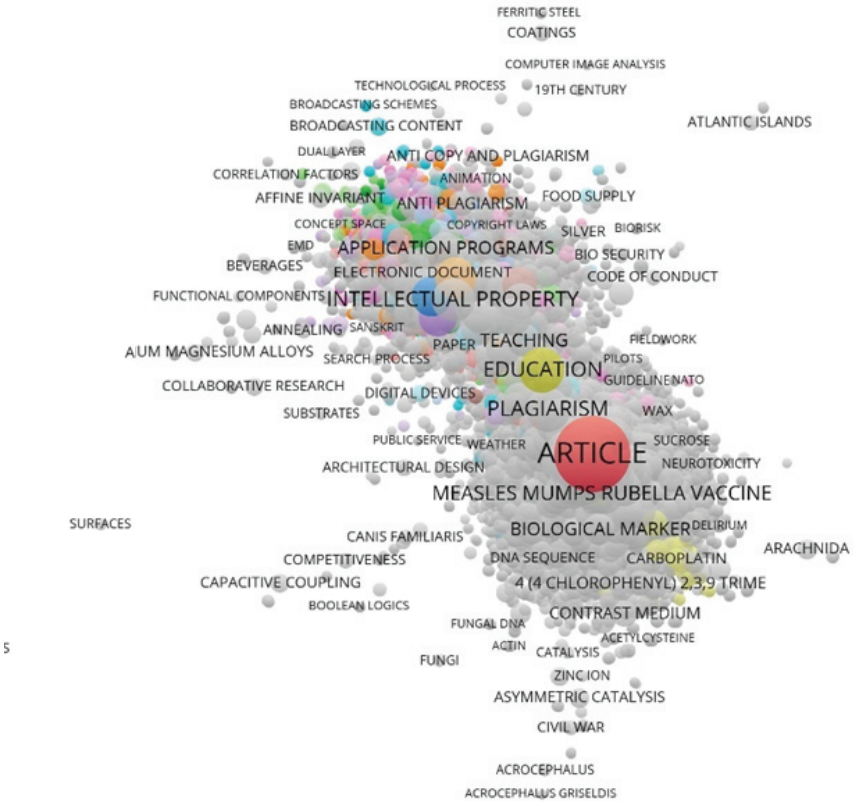

Figure 4. Keyword map of global literature using VOS viewer.

using the SNA tools, the whole community relationship in terms of the whole network level and the individual actor level can be derived. SNA tools can describe the basic assumptions, goals, and explanatory mechanisms prevalent in the field[17-18]. This section of the article will find the relationship among the keywords of the subject both individual actor and the whole keywords dynamics in this filed both from India and from the global literature. The study is also further trying to analyse the top occurring keywords based on their centrality network from the globally published literature as well as from India.

The whole network of global literature has 15,329 node and 61,566 edge. Here nodes are the keywords and edges are the connections between them. The whole network has a degree of 8.033. The average clustering coefficient is 0.285, Network diameter is 10, and the average pathlength is 3.766. It means from one point of a network to its furthest point is ten units' length, and it requires 3.76 average paths to reach from one actor to other. The whole keyword network diagram is as shown in Fig. 5. The network has formed 649 cluster. The biggest cluster has 170 keyword items. There are about 250 cluster with more than ten keyword.

Centrality measures are the most important and widely used tools in SNA. These indicators give fair idea of centrally located and important actors within the sample. It also shows the potential flow of knowledge and communication between the actors. The widely used centrality measure is the degree centrality. It shows the number of occurrences of an actor in a network. It is measured by counting the number of direct links from an actor to others in the network. So, the maximum degree for an actor in a network means they are the powerful, influential and the connector of the network[19]. Here in the keyword mapping context, in the simplest term the greater number of occurrences of an actor means it is the prominent and powerful keyword. Table 1 shows the degree centrality measures of keywords in global as well as in Indian context.

4.7 Keyword Analysis for Indian Literature

Keywords from Indian literature has 1,612 node and 2,982 edge. The whole network has an average degree of 3.684, and average weighted degree is 4.795. The network diameter is 9, and it has average clustering coefficient 0.243, the average path length is 4.047. So, it is obvious that the network of the keyword is small compared to the global network. The whole network is as seen in Fig. 5.

Keywords derived from the Indian published literature has formed 164 clusters, and the largest cluster has 58 items. There are 74 clusters with more than ten items.

5. CORE JOURNALS

The subject now becomes increasingly growing and broad over the years. So, the identification of core journals in this field can be useful[20] for library and information science professionals (LIS) in the journal selection process. It is perhaps helpful for research scholars for selecting the journal for finding precise information and as the research outlet of their research works.

Table 1. Top 10 keywords with high degree centrality measures from Global vs Indian publications

Global literature		Indian literature	
Keywords	**Degree centrality measures**	**Keywords**	**Degree centrality measures**
Article	2241	Article	190
Scientific misconduct	848	Human	106
Human	1152	Ethics	68
Editorial	780	Editorial	76
Ethics	798	Scientific misconduct	53
Adult	720	Publishing	53
Publishing	570	Codes (symbols)	88
Humans	700	Artificial intelligence	109
Conflict of interest	468	Author	75
Publication	437	Intellectual property	79

Figure 5. Network structure of keyword for Indian research literature.

The scattering of articles shows that the distribution of articles does not follow a typical 'Bradford' distribution. The articles are scattered in a number of varied sources. As the subject is in the growing stage, it is yet to form a core set of journals. The highest number of research articles have been published in the *Nature* (320 articles) followed by *Science and Engineering Ethics* (239) and *Science* (206). From India, the highest number of publications are in the journal *Current Science* (23 articles). Table 2 shows the top 10 journal with about 50 publication.

The subject under investigation is a multidisciplinary research area encompassing a diverse field of knowledge. Hence a large number of articles published in the multidisciplinary journal. For example, both globally as well as from India a large number of publications comes out from the multidisciplinary journals like nature, science, current science and so on.

Also, the important observation from the Indian context is a good number of publications from India were published in Indian journals.

6. MOST PRODUCTIVE INSTITUTIONS

It is important to analyse institutional affiliation to map the productive institutions. Table 3 shows the top productive institutions. Globally Monash University, Australia is the most productive institution with 62 publications, followed by Universitat Politècnica de València (UPV) in Spain, and the

Table 2. Top 10 journals of publications from globally and from India

Journals from outside India		Journals from India	
Source	**Publications**	**Source**	**Publications**
Nature	320	*Current Science*	23
Science and Engineering Ethics	239	*CEUR Workshop Proceedings*	12
Science	206	*Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*	9
NIH Guide for Grants and Contracts Online	170	*Economic and Political Weekly*	7
BMJ Clinical Research Ed	166	*Indian Journal of Dental Research*	6
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	158	*Indian Journal of Medical Ethics*	6
CEUR Workshop Proceedings	143	*Journal of Indian Society of Pedodontics and Preventive Dentistry*	6
Accountability in Research	123	*Medical Journal Armed Forces India*	6
Lancet	95	*ACM International Conference Proceeding Series*	5
Journal of Academic Ethics	92	*Advances in Intelligent Systems and Computing*	5
		IEEE	5
		5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services ETTLIS 2018	5
		Journal of Pharmacology and Pharmacotherapeutics	5
		Procedia Computer Science	5
		Science and Engineering Ethics	5

Table 3. Top 10 productive institutions with number of publications

Institutions from outside India		Institutions from India	
Institutions	**Publications**	**Institutions**	**Publications**
Monash University	62	Amrita Vishwa Vidyapeetham, Bangalore	23
Universitat Politècnica de València	61	All India Institute of Medical Sciences, New Delhi	9
University of Toronto	59	Jawaharlal Institute of Postgraduate Medical Education & Research	8
University of Oxford	59	Indian Academy of Sciences	8
The University of Sydney	49	Panjab University	8
University of Michigan, Ann Arbor	49	Banaras Hindu University	7
Deem Corporation	46		
University of Tehran	44	Manipal Academy of Higher Education	6
Harvard Medical School	43	Lovely Professional University	6
University of Manchester	43	Postgraduate Institute of Medical Education and Research Motilal Nehru National Institute of Technology Armed Forces Medical College Anna University Sanjay Gandhi Postgraduate Institute of Medical Sciences Lucknow Ajay Kumar Garg Engineering College	5 5 5 5 5

University of Toronto in Canada. From India, Amrita Vishwa Vidyapeetham, Bangalore is the most productive institution with 23 publications, followed by the All India Institute of Medical Sciences (AIIMS) in New Delhi, and Jawaharlal Institute of Postgraduate Medical Education & Research (JIPMER) in Puducherry.

7. MOST PRODUCTIVE COUNTRIES

Table 4 enlists the top productive countries with more than 100 publications. The United States is the most productive country with 2,928 Scopus record, i.e., about 26 percent of global literature. India's contribution is about 459 record, which constitutes about four percent of global literature. India's contribution is ranked 4th after Australia. Nevertheless, India is placed above Germany and China. Interestingly, China is the second major publisher of scientific literature globally; but Chinese contribution in this area is comparatively lower than India and other top countries with significant S&T publications.

Table 4. Top 10 productive countries

Country	Publications	Per centage share in global literature
United States	2,928	25.87
United Kingdom	1,105	9.76
Australia	464	4.10
India	459	4.06
Germany	449	3.97
China	390	3.45
Canada	377	3.33
Spain	282	2.49
Netherlands	214	1.89
France	192	1.70

8. CONCLUDING REMARKS

The paper is an attempt to map the global literature related to academic integrity and plagiarism through the scholarly publication data available from the globally available indexing and citation database. The Scopus database was chosen because of its wide coverage of science and technology literature. The period of investigation was from 2000 to 2018. The publications from 2000 were chosen because of the worldwide proliferation of higher education and also the exponential growth of scholarly literature. In this context, this is a timely study to map the growth and dynamics of the subject. The study observed that there is a growth of scholarly literature in recent years. Along with the growth of publications globally, Indian scholarly literature on this field is exponential. The recently adopted policy documents by the Government of India will perhaps boost further literature growth in this nascent area.

REFERENCES

1. Macfarlane, B.; Zhang, J. & Pun, A. Academic integrity: A review of the literature. *Stud. Higher Educ.,* 2014 **39**(2), 339-358. doi: 10.1080/03075079.2012.709495

2. Korenman, S.; Berk, R.; Wenger, N. & Lew, V. Evaluation of the research norms of scientists and administrators responsible for academic research integrity. *J. Am. Med. Assoc.,* 1998, **279**(1), 41-47.
doi: 10.1001/jama.279.1.41.
3. Kadam, D. Academic integrity and plagiarism: The new regulations in India. *Indian J. Plast. Surg.,* 2018 **51**(2), 109-110.
doi: 10.4103/ijps.IJPS_208_18.
4 KAUST Library, Saudi Arabia 2018. Plagiarism and how to avoid it: Avoiding plagiarism. https://libguides.kaust.edu.sa/plagiarism/avoid (accessed on 15 February 2019).
5. Bagla, P. India creates unique tiered system to punish plagiarism. *Science,* 2018. https://www.sciencemag.org/news/2018/04/india-creates-unique-tiered-system-punish-plagiarism (accessed on 10 February 2019).
doi: 10.1126/science.aat8300.
6. Chandra, S. India cracks down on plagiarism at universities: But some researchers say new rules don't go far enough. *Nature*. https://www.nature.com/articles/d41586-018-05924-1 (accessed on 10 February 2019).
doi: 10.1038/d41586-018-05924-1.
7. UGC, India. University Grants Commission (Promotion of academic integrity and prevention of plagiarism in higher educational institutions) Regulations, 2018. The Gazette of India: Extraordinary [Part III, Sec. 4], 31st July 2018. New Delhi: University Grants Commission.
8. Scopus: Content coverage guide. 2017. https://bit.ly/2BN3YHu (accessed on 15 February 2019).
9. Bastian, Mathieu; Sebastien, Heymann & Mathieu, Jacomy. "Gephi: An open source software for exploring and manipulating networks." Third international AAAI conference on weblogs and social media. 2009. https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009. Accessed on 31 January 2019.
10. Van Eck, N. & Waltman, L. Software survey: VOSviewer, A computer program for bibliometric mapping. *Scientometrics,* 2009, **84**(2), 523-538.
doi: 10.1007/s11192-009-0146-3.
11. Borgatti, S.P.; Everett, M.G. & Freeman, L.C. Ucinet for windows: Software for social network analysis. Harvard, MA: Analytic technologies. 2002
12.. He, Q. Knowledge discovery through co-word analysis, 1999. https://bit.ly/2Six9Gw (accessed on 15 February 2019).
13. Chen, X.; Chen, J.; Wu, D.; Xie, Y. & Li, J. Mapping the research trends by co-word analysis based on keywords from funded project. *Procedia Comp. Sci.,* 2016, **91**, 547-555.
doi: 10.1016/j.procs.2016.07.140.
14. Ding, Y.; Chowdhury, G.G., & Foo, S. Bibliometric cartography of information retrieval research by using co-word analysis. *Inf. Process. Manage.,* 2001, **37**(6), 817-842.
doi: 10.1016/S0306-4573(00)00051-0.
15. Van Raan, A.F.J. & Tijssen, R.J.W. The neural net of neural network research. *Scientometrics,* 1993, **26**(1), 169-192.
doi: 10.1007/BF02016799.
16. Scott, J. Social network analysis. *Sociology,* 1988, **22**(1), 109-127.
doi: 10.1177/0038038588022001007.
17. Borgatti, S.P.; Mehra, A.; Brass, D.J. & Labianca, G. Network analysis in the social sciences. *Science,* 2009, **323**(5916), 892-895.
doi: 10.1126/science.1165821.
18. Wasserman, S. & Faust, K. Social network analysis: Methods and applications (Vol. 8). Cambridge: Cambridge University Press. 1994
19. Hanneman, Robert A. & Mark Riddle. Introduction to social network methods. Riverside, CA: University of California, Riverside, 2005, (published in digital form at http://faculty.ucr.edu/~hanneman/ accessed on 15 February 2019)
20. Patra, Swapan Kumar & Chand, Prakash. Biotechnology research profile of India. *Scientometrics*, 2005 **63**(3), 583-597.
doi: 10.1007/s11192-005-0229-8.

CONTRIBUTORS

Dr Swapan Kumar Patra did his PhD in Science Policy from Jawaharlal Nehru University, New Delhi, India. Presently a Post-Doctoral Researcher at Tshwane University of Technology, Pretoria, Republic of South Africa. His research interest is in globalisation of R&D with particular reference to developing countries, science technology indicators, scientometrics, innovation systems, open innovation and social network analysis.
He contributed in this study by doing literature review, framing research objectives, data analysis, and writing.

Dr Anup Kumar Das received his PhD from Jadavpur University, Kolkata, in 2009. He is an alumnus of CODATA-RDA Data Science International Training School. Presently working in the Centre for Studies in Science Policy at Jawaharlal Nehru University, India. His research interests revolve around open access, open educational resources (OER), open research data, information policies, knowledge societies, scientometrics, and altmetrics. He is the Co-Convener of 'Open Access India' – an advocacy group for promotion of open access, open science and OER.
In this study, he conceptualized the paper, collected the data from Scopus databse, and wrote jointly with the co author.