

EXAMINATION OF COMPUTER SCIENCE STUDIES IN INDIA AND THE UNITED STATES USING SCOPUS

Ms. Pournima Jangle, Mr. Avinash Bhagat, Faculty, Dept. of Computer Science Tilak Maharashtra Vidyapeeth, Pune-37

Abstract

In recent times, there has been a noticeable surge in the inclination of Indian authors towards the domain of research impact analysis, with a specific focus on Indian educational institutions across several disciplines. This study undertakes a comprehensive analysis of research impact, specifically examining a set of Indian institutes in contrast to their counterparts in the United States, with a specific focus on the field of computer science. In order to accomplish this project, a thorough dataset going up to January 2023 has been obtained from Scopus. The collected data is subsequently submitted to an extensive analysis that utilizes a diverse range of quantitative and qualitative citation measures. Python, a highly adaptable programming language widely recognized for its effectiveness in the fields of data analysis and visualization, is employed to facilitate the aforementioned analysis. The main aim of this study is to provide light on the overall research effect of specific Indian Institutes and United States in the field of Computer Science. Through the examination of diverse citation metrics, this research aims to offer significant perspectives on the international reputation and impact of these establishments, therefore enhancing comprehension of the scholarly environment in the field of Computer Science in India and the United States.

Keywords: g-index, Scopus, h-index, Citations, Analysis.

Introduction

There has been a recent trend in Indian universities, where the administration is attempting to shift the academic climate in favour of research. It becomes much more crucial when academic rating in NIRF is replaced with criteria for credentials such as research article, conference, and citation (National Institutional Ranking Framework, MHRD). When someone wants to grow up an Indian university as a research institution in the specified field, an issue with how to analyze the current trend of Indian universities occurs. When data must be retrieved from Scopus, which has more than 36,000 journals, it becomes much more important (www.scopus.com). At the same time, Scopus has more than 40,000 conferences listed. Numerous conferences used to index the publications that various journals sent out from Scopus based on their quality each year. It demonstrates how important it is for data analysis researchers to measure trends at any university based on both their quality and quantity. As a result, the author of this study concentrates on the computer science research sector to analyze Indian Institutes and the USA.

When such a vast amount of data is indexed in Scopus, one of the most important responsibilities is the quantitative and qualitative analysis of the research impact of any institution. A measure of quantity may be the total number of articles from all institutes that have been published in conferences or journals with the Scopus index. However, overall citations, journal publications, most cited papers; citations from outside of universities, papers from the top institutes, top journal articles and technical papers can be considered as reliable indications of the quality of publications. The manipulation of citations and papers is a widely acknowledged phenomenon. As a result, citation metrics like the h-index (Hirsch, 2005) and g-index (Egghe, 2006) are used to gauge the quality and quantity of the effect of research. The h-index evaluates N papers with at least N citations in terms of both quantity and quality. However, it does not account for influential research or papers that have received a lot of citations over time. It has been suggested that the h-index is unable to provide any information concerning the popular or extremely significant research that has been carried out by any institute. To do this, an overview is provided by the g-index. Due to this, some scholars (Prathap & Gupta, 2009; Gupta, 2010; Singh & Sinha, 2018; Singh & Singh, 2019; Singh, 2020; Elsevier, 2016; Prathap, 2013; Uddin & Singh, 2015; Parmar & Siwach, 2016; Banshal et al., 2017) attempted to rank Indian institutions using Scopus data sets (Singh, 2020; Elsevier, 2016; Prathap, 2013; Uddin &

Singh, 2015; Parmar & Siwach, 2016; Banshal et al., 2017) based on total publications, citations per paper (Prathap & Gupta, 2009; Gupta, 2010; Singh & Sinha, 2018; Singh & Singh, 2019), and other factors (Singh, 2020). Singh and Chandresh have detailed the thorough bibliometric analysis of Indian Institutes in the field of computer science based on the provided factors. It has been noted that the ranking conducted in (Singh & Singh, 2019) November 2018 and the NIRF (National Institutional Ranking Framework, MHRD) April, 2019 announcement by the Indian Government are roughly equivalent. To ascertain the research tendencies of Indian Institutes in the field of computer science, the authors acknowledged the necessity of comparing the obtained results with those of institutes in the United States. This study aimed to quantitatively assess the research impact of two nations by considering various metrics, including total computer science publications, total citations, h-index, and g-index. Specific institutes were selected to examine this matter. The objective is to provide a comprehensive summary of the collaborative research endeavours in computer science between Indian colleges and institutions in the United States, with the aim of facilitating future breakthroughs in the field. The objective of this study is to conduct an analysis of the research trends in both countries, utilizing the factors that have been presented. In order to enhance its ranking, any institution might prioritize these aspects. Other domain experts will find it useful to analyze their rating at the same time.

Background

The following definitions are provided in this section in order to help readers comprehend the analysis of the current paper:

Scopus

Elsevier launched Scopus, greatest database of abstracts and citations from scholarly journals, in November 2004 (Elsevier, 2019; 2017). It includes a number of tools for analyzing, following, and data visualization for over 36,000 publications and 40,000 conference papers from over 11,000 publishers. It can be regarded as one of the most well-known databases to analyze the influence of any institutes' research. Total publications and total citations are the two most used ways for quantitatively evaluating research output. The most thorough overview of global research output in the fields of technology, science, social science, medicine, humanities, and the arts is provided by Scopus. Total publications and total citations are the two most used ways for quantitatively evaluating research output. There has always been some disagreement over how to judge the quality of scientific findings. The measures employed in this study for measuring research output and their importance is outlined below.

Number of Publications (P)

The overall amount of research papers a certain author has written are reflected in their total number of publications. It can be used to gauge how much research is produced by a certain university. Conference and journal counts are treated similarly by Scopus for the specified paper. Currently, there are 11 different document types listed in the Scopus database, including chapter, editorial, press article, article, book, erratum, review, letter, report, note, short survey, and conference paper. The two main categories of this study are conference papers and articles (including articles in the press). This statistic just offers quantity, not quality.

Citations (C)

The number of citations indicates how frequently a paper is cited. It denotes high calibre or current trend study that is acknowledged by other academics. Although a publication's total number of citations provides information about the effect of the study, it has never been clear how many citations actually indicate high-calibre research. The citation can be changed at the same time. Self-citations, citations from sources outside of universities, citations from prestigious journals, citations from review articles, and citations from technical publications are now another study concern as addressed in (Singh, 2020). This essay solely focuses on the idea that any citation counts as one.

Citation per paper (CP) is a ratio of how many times an author has been cited to how many times they have published. When a researcher's overall influence is considered rather than just the number of citations, this term is used.

h-index

The h-index (Hirsch, 2005) measures how many times each of an author's works has been cited by other works in the field. It gives insight into both the number and calibre of an academic's research output. The fact that the h-set does not take into consideration the contribution of research papers outside of it is one of the shortcomings of the h-index. At the same time, it offers no data on the most popular or frequently cited papers over time.

g-index

The greatest value of g where at least g² of the author's articles have been referenced in total is used to define the g-index. It is an improvement on the h-index since it takes into account works that have received few citations and counterbalances them with the citations received by works that have received many references. If the h-index represents the number of high-quality research publications, then the g-index can be looked of as the impact of these studies.

Motivation for this Study

In comparison to American institutes, this study gives an overall analysis of computer science publications in journals, conferences, citations/paper, and h- and g-indices from Indian institutes. The research in this field is limited to ranking or contrasting Indian universities with one another. Finding the disparity between the research output of top American institutes and that of Indian institutes in the field of computer science is the driving force behind this study. It provides us with a picture of India's position on the global research stage.

Objective of the Study

The aim of this study is to assess the quantity and quality of Indian Institutes in comparison to those in the United States, based on citation data from publications in Computer Science linked with a selected academic institution.

Methodology

Following is a list of the steps involved in gathering data:

- 1) Scopus offers a csv download option for the citation data for each relevant academic institution.
- 2) The institutions that had over 20,000 articles were downloaded by arranging them in descending and ascending order based on their publication year.
- 3) The downloaded data cleaned by adding the necessary columns to the csv file and fixing any missing entries or double quotes.
- 4) The data set was ultimately consolidated into a single.csv file, encompassing the data from the 23 selected institutes.
- 5) The next stage was to determine the various metrics needed for the comparison of academic institutes.
- 6) Python programs were created utilizing the numpy and pandas libraries to compute citation metrics and other parameters.
- 7) The dataset's EID and Cited by fields are of particular relevance. The EID used as the primary key because it is the only way to identify an academic work.
- 8) The amount of EIDs that are present in an institute's dataset constitutes the entire number of publications.
- 9) The "Cited by" entry indicates how many times each publication in the dataset has been cited. All of the numbers in this field are added to determine an institution's overall citation count.
- 10) The procedures provided in Figure 2 and 1 was used to calculate the g-index and h-index of institutes respectively.

```
C ← Array of citations for each publication
C ← Sort_descending(C)
h ← 0
while (h < C.length AND C[h] > h):
    h ← h+1
```

Figure 1

```
C ← Array of citations for each publication
C ← Sort_descending(C)
g ← 0
g_citation ← 0
for c in C:
    if g >= C.length:
        break
    if g_citation >= g * g
        g_citation += c
        g += 1
```

Figure 2

Results

Tables 1 to 4 show the results from the data sets that were extracted together with the outputs they produced. Table 1 lists the total number of computer science publications and citations from Indian Institutes and the USA. While Table 2 displays their citations per manuscript,. h-index and g-index are shown in tables 3 and 4, respectively.

Table 1 the total number of computer science publications and citations from Indian Institutes and the USA

	Indian	Citation s	Publicati ons	USA	Citatio ns	Publicatio ns
1	IIT, Hyderabad	5328	1151	Duke University (DU) Durham	160693	9470
2	IIT, Indore	6372	958	University of Chicago (UChicago), Chicago	170875	6712
3	NIT, Surat	8200	1923	Johns Hopkins University Baltimore	210980	9776
4	University of Calcutta	9853	2156	Yale University, New Haven	219104	7328
5	University of Delhi (DU)	12040	2254	Northwestern University Evanston	244771	11300
6	TIFR, Mumbai	12694	1340	Harvard University Cambridge	258328	9159
7	BITS, Pilani	12825	2175	University of Pennsylvania Philadelphia	291660	12841
8	Thapar University, Patiala	13699	2213	Columbia University	318527	13743
9	NIT, Tiruchirapalli	14471	2145	New York University (NYU)	337348	3089
10	NIT, Rourkela	15079	2804	California Institute of Technology (Caltech or CIT) Pasadena	380944	14507
11	IIIT, Hyderabad	16731	2475	Purdue University, Indiana	395368	22119
12	VIT, Vellore	16949	4592	Princeton University Princeton	410401	13334
13	IIT, Guwahati	18780	3290	Cornell University, Ithaca	447040	15523
14	Anna University, Chennai	27169	6418	University of Washington Seattle	451069	17118
15	IIT Roorkee	33150	4331	University of California, San Diego	505845	17871
16	Jadavpur University (JU), Kolkata	47175	5524	University of Michigan, Ann Arbor	534417	21250
17	IIT, Madras	55800	6236	University of California Los Angeles	561806	17944
18	IIT, Bombay	57842	6490	Georgia Institute of Technology, Atlanta	581186	27964
19	Indian Statistical Institute (ISI), Kolkata	59361	4161	University of Illinois at Urbana- Champaign	733518	27390
20	IIT, Kanpur	74329	4714	Carnegie Mellon University Pittsburgh	971704	33515
21	IIT, Delhi	76470	6636	Massachusetts Institute of Technology (MIT) Cambridge	1364472	30463
22	IIT, Kharagpur	78091	7932	University of California Berkeley	1393346	30257
23	IISc, Bangluru	88430	7693	Stanford University, Stanford	1468632	29203
Mean		33079.913	3896.1304		539653.65	17472.87
Total		682747	81679		1101863	371619

				88	
BITS: Birla Institute of Technology and Science; TIFR: Tata Institute of Fundamental Research; IISc: Indian Institute of Science; IIT: Indian Institute of Technology; NIT: National Institute of Technology; IIIT: International Institute of Information Technology; VIT: Vellore Institute of Technology					

The average publishing rate for Indian Institutes i.e. 15516 is not achieved by a single institute, as shown in Table 1. It was also observed citations of USA institutes were higher as compare to India. There is huge difference reported in number of citations as well as Publications. India is far behind in citations count.

Table 2 Citations per paper in the USA and India's Institutes

	Indian	(CP)	USA	(CP)
1	Vellore Institute of Technology (VIT), Vellore	3.691	Duke University (DU), Durham	16.97
2	Anna University, Chennai	4.233	Purdue University, Indiana	17.87
3	National Institute of Technology Surat	4.264	Georgia Institute of Technology, Atlanta	20.78
4	University of Calcutta	4.57	Johns Hopkins University, Baltimore	21.58
5	IIT, Hyderabad	4.629	Northwestern University, Evanston	21.66
6	University of Delhi	5.342	University of Pennsylvania, Philadelphia	22.71
7	NIT, Rourkela	5.378	Columbia University, New York City	23.18
8	IIT, Guwahati	5.708	University of Michigan, Ann Arbor	25.15
9	BITS, Pilani	5.897	University of Chicago (UChicago), Chicago	25.46
10	Thapar University, Patiala	6.19	California Institute of Technology (Caltech or CIT), Pasadena	26.26
11	IIT, Indore	6.651	University of Washington, Seattle	26.35
12	NIT, Tiruchirapalli	6.746	University of Illinois at Urbana- Champaign	26.78
13	IIIT, Hyderabad	6.76	Harvard University, Cambridge	28.2
14	IIT, Roorkee	7.654	University of California San Diego	28.31
15	Jadavpur University (JU), Kolkata	8.54	Cornell University, Ithaca	28.8
16	IIT, Bombay	8.912	Carnegie Mellon University, Pittsburgh	28.99
17	IIT, Madras	8.948	Yale University, New Haven	29.9
18	TIFR, Mumbai	9.473	Princeton University, Princeton	30.78
19	IIT, Kharagpur	9.845	University of California, Los Angeles	31.31
20	IISc, Bangluru	11.495	Massachusetts Institute of Technology (MIT), Cambridge	44.79
21	IIT, Delhi	11.524	University of California, Berkeley	46.05
22	Indian Statistical Institute, Kolkata	14.266	Stanford University, Stanford	50.29
23	IIT, Kanpur	15.768	New York University (NYU)	51.99
	Mean	7.67		29.31

Based on the data presented in Table 2, it can be observed that the mean number of citations per manuscript in the United States is almost four times greater than that in India. Table 1 demonstrates a notable disparity in the overall number of citations between the United States and India, with the former exhibiting a magnitude 15 times more than the latter. However, it is worth noting that the average number of citations in the United States only exhibits a slight increase in comparison to India. The difference ratio decreases from 4 to 3 when we compare the two best institutions in each nation. It demonstrates that there are very few high-calibre research institutions in India, which lowers the country's average for citations per publication. It implies that the majority of research conducted by Indian institutes is not focused on current trends or that they have not published as much significant research as the United States. At the same time, in India any article with more than seven citations is regarded as an averagely good work, whereas in the USA it is 29. It is clear that none of the Indian institutes exceeds the standard for citations in American institutes' papers. It

implies that Indian institutions might not work on more impactful or current developments. Table 3 displays their analysis using the h-index to measure it.

Table 3 analysis using the h-index

	Indian	h-index	USA	h-index
1	IIT, Hyderabad	24	Duke University (DU), Durham	145
2	NIT, Surat	30	University of Chicago, Chicago	150
3	IIT, Indore	32	Johns Hopkins University, Baltimore	163
4	University of Calcutta	37	Yale University, New Haven	176
5	University of Delhi (DU)	39	Northwestern University, Evanston	182
6	BITS, Pilani	42	Harvard University, Cambridge	189
7	NIT, Rourkela	43	University of Pennsylvania, Philadelphia	196
8	VIT, Vellore	44	Columbia University	201
9	TIFR, Mumbai	47	New York University (NYU)	209
10	Thapar University, Patiala	47	Purdue University, Indiana	209
11	IIIT, Hyderabad	48	California Institute of Technology (Caltech or CIT), Pasadena	217
12	NIT, Tiruchirapalli	52	Princeton University, Princeton	249
13	IIT, Guwahati	52	University of Michigan, Ann Arbor	252
14	Anna University, Chennai	52	Georgia Institute of Technology, Atlanta	254
15	IIT, Roorkee	68	University of Washington, Seattle	255
16	Jadavpur University, Kolkata	83	Cornell University, Ithaca	256
17	IIT, Bombay	88	University of California, San Diego	259
18	IIT, Madras	90	University of California, Los Angeles	267
19	IIT, Kanpur	90	University of Illinois at Urbana-Champaign	295
20	Indian Statistical Institute (ISI), Kolkata	95	Carnegie Mellon University, Pittsburgh	347
21	IIT, Kharagpur	99	Massachusetts Institute of Technology, Cambridge	385
22	IIT, Delhi	99	University of California Berkeley	392
23	IISc, Bangluru	104	Stanford University, Stanford	412
Mean		61.08		246.08

According to Table 3, the average h-index for the USA is around four times greater than for India. The h-index ratio between the highest-ranking institutes in both countries is approx. 4 (3.96), however the h-index of the lowest-ranking institute in the United States is around six times bigger than that of India. This observation implies that India experiences a more pronounced decrease in h-index compared to the United States. It supports the findings gained from Table 2 by confirming that only a small number of Indian institutions do both high-quality and quantity research, such as IIT Delhi, IISc Bangluru, ISI, and IIT Kharagpur. It does not, however, state that any Indian institutions conducted any significant study during that time. Table 4 offers their comparison analysis using the g-index to address this problem.

Table 4 comparison analysis using the g-index

	Indian	g-index	USA	g-index
1	IIT, Hyderabad	36	Duke University (DU), Durham	236
2	NIT, Surat	51	Northwestern University, Evanston	300
3	Indian Institute of Technology Indore	51	Johns Hopkins University, Baltimore	304
4	University of Calcutta	55	University of Chicago (UChicago)	309
5	University of Delhi (DU)	56	University of Pennsylvania, Philadelphia	330
6	NIT, Rourkela	63	Columbia University	334
7	VIT, Vellore	66	Purdue University, Indiana	344
8	Thapar University, Patiala	69	Yale University, New Haven	348
9	NIT, Tiruchirapalli	71	Harvard University, Cambridge	359
10	BITS, Pilani	75	Princeton University, Princeton	416
11	IIT, Guwahati	79	University of Washington, Seattle	422
12	TIFR, Mumbai	80	Georgia Institute of Technology, Atlanta	447
13	IIT, Hyderabad	81	New York University (NYU)	454
14	Anna University, Chennai	90	California Institute of Technology (Caltech or CIT), Pasadena	458
15	IIT, Roorkee	103	Cornell University, Ithaca	466
16	IIT, Madras	139	University of California San Diego	486
17	Jadavpur University, Kolkata	143	University of Michigan, Ann Arbor	492
18	IIT, Bombay	151	University of Illinois at Urbana-Champaign	524
19	IIT, Kharagpur	158	University of California Los Angeles	530
20	Indian Statistical Institute (ISI), Kolkata	172	Carnegie Mellon University, Pittsburgh	575
21	IIT, Delhi	173	Massachusetts Institute of Technology (MIT), Cambridge	706
22	IISc, Bangluru	198	Stanford University, Stanford	745
23	IIT, Kanpur	216	University of California, Berkeley	779
	Mean	103.3		450.6

IIT Kanpur is shown to have a g-index of 216 in Table 4. This suggests that IIT Kanpur has generated significant research that has garnered numerous citations throughout the course of time. One of the Indian Institutes competes with USA Institutes in this parameter. The current article can help the management of the government, corporate organizations, and other institutions raise the quantity and quality rankings of their universities in this way. The following highlights from the current study may be helpful when making decisions: During the course of the academic year, publish some of the most influential research publications.

Any institution or administration that wishes to boost its university rating in the computer science field can benefit from the aforementioned analyses. It will be useful in helping them make several decisions based on their needs and capabilities. The primary focus of the concurrent work does not revolve around time-based publishing or citation analysis. The analysis does not prioritize characteristics such as self-citations, citations between conferences, citations from external sources, or other forms of exhaustiveness. The author intends to focus on the development and implementation of more measures in the foreseeable future in order to effectively tackle this issue.

Conclusion

This study compares the quality and number of computer science articles across Indian and American institutions using the Scopus database. When compared to Indian Institutes, it can be shown that each USA institute publishes more than six times as often. Indian institutes are considerably behind American institutes in terms of citation per publication and h-index. It suggests that, in contrast to

Indian Institutes, American Institutes published more high-quality and quantity papers on current trends. It demonstrates that Indian Institutes publish 4 times less than USA Institutes within the same time period. According to Table 1, the Indian Institutes publish less and of lower quality. The Tables 2 and 3 show that no Indian institutes even come close to the average number of citations per manuscript or the h-index of US institutions. It signifies that when compared to the USA, Indian Institutes fall short in terms of both quantity and quality of computer science papers. At the same time, the majority of research may or may not be outdated or unhelpful. Because of this, they averaged just 10 citations per paper. The g-index displayed in Table 4 at the same time shows that IIT Kanpur's index nearly reaches Duke University. When compared to other Indian Institutes, it indicates that the IIT Kanpur has conducted some of the most influential or cutting-edge computer science research over the years. The administration of numerous institutes will find this material useful in improving their indexing rankings. Additionally, it will be useful for computer science students applying for admission.

References

1. B. V. Elsevier, Academic Research. *Elsevier Developers*, https://dev.elsevier.com/academic_research_scopus.html, Retrieved May 01, 2019.
2. Banshal, S. K., Singh, V. K., Basu, A., & Muhuri, P. K. (2017). Research performance of Indian institutes of technology. *Current Science*, 923-932.
3. Egghe, L. (2006). An improvement of the h-index: The g-index. *ISSI newsletter*, 2(1), 8-9.
4. Elsevier, S. (2016). Scopus content coverage guide. *Amsterdam: Elsevier BV*.
5. Gupta, B. M. (2010). Ranking and performance of Indian Universities, based on publication and citation data. *Indian Journal of Science and Technology*, 3(7), 837-843.
6. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102(46), 16569-16572.
7. National Institutional Ranking Framework MHRD <https://www.nirfindia.org/2019/EngineeringRanking.html>, Retrieved May 01, 2019.
8. Parmar, S., & Siwach, A. (2016). Indian Research output in Computer Science during 2004-2013: a bibliometric analysis. *International Journal of Digital Library Services*, 6(2), 20-31.
9. Prathap, G., & Gupta, B. M. (2009). Ranking of Indian engineering and technological institutes for their research performance during 1999–2008. *Current Science*, 97(3), 304-306.
10. Prathap, G. (2013). Benchmarking research performance of the IITs using " Web of Science" and " Scopus" bibliometric databases. *Current Science*, 1134-1138.
11. Scopus, www.scopus.com
12. Singh, P. K., & Sinha, S. (2018). Analyzing Indian Universities using Scopus Dataset. In *Proceedings of International Conference on Intelligent Computing and Sustainable System (ICICSS-2018)*.
13. Singh, P. K., & Singh, C. K. (2019, February). Bibliometric study of Indian Institutes of Technology in computer science. In *2019 Amity International Conference on Artificial Intelligence (AICAI)* (pp. 384-393). IEEE.
14. Singh, P. K. (2020). Multi-granular-based n-valued neutrosophic context analysis. *Granular Computing*, 5(3), 287-301.
15. Uddin, A., & Singh, V. K. (2015). A quantity–quality composite ranking of Indian institutions in CS research. *IETE Technical Review*, 32(4), 273-283.