Quest for Ranking Excellence: Impact Study of Research Metrics

Sheeba Pakkan*, Christopher Sudhakar, Shubham Tripathi and Mahabaleshwar Rao

Manipal Academy of Higher Education, Manipal - 576 104, India
*E-mail: sheeba.pakkan@manipal.edu

ABSTRACT

It is a general trend that higher education institutions must undergo one or more ranking agencies evaluation to get reputed in the field of education. Institutions must have a place in the ranking to attract students, faculty and even for successful foreign student exchange programs. Prominence in research is an important parameter to scale the eminence of an institution. A quantum of quality publications is one of the factors judging the prevailing research environment of an institution. While doing an institutional level evaluation on research output, we always go for quantitative measures like total publication, citation and h-index. In the present scenario, the institution must set to have a strategy in achieving ambitious scientific goals. This paper presents some innovative analysis of research data by choosing valuable metrics. The samples selected for the study are the top-ranked institutions in India. The same approach can be applied to any institutions in the world for evaluation and excellence. We have done qualitative-quantitative analysis using different parameters of research output to explain the importance of various metrics. The present study has identified metrics which can be considered seriously to achieve better performance in research output. It is a fused index metrics which explains along with Scholarly Output, the importance of Cited and Uncited output. It is a fused index metrics which explains along with Scholarly Output, the importance of Cited and Uncited publications and different types of Collaborations and its impact on the overall performance of an institution.

Keywords: Impact study; Bibliometric; University ranking; Big data analytics; Statistical analysis

1. INTRODUCTION

Challenging changes are happening these days in the field of higher education. Across the world, institutions are looking for ranked universities for exchange of students as well as research collaborations. Nowadays, universities are considered not only as a place for teaching but also as an institution where research is given equal or more importance. World University Ranking is bringing different parameters year on year to evaluate and rank the universities. Indian universities are trying to manage and grab a suitable place on top among the world-leading universities but unfortunately, fail to reach out to the top 200 in the Times Higher Education Ranking 2020. In India, at present, about 900 universities are providing higher education to fulfill the educational needs of the country. These days’ different performance ranking exercises are being employed among the higher education institutions in India, like the National Institute of Ranking Framework (NIRF) and National Assessment and Accreditation Council (NAAC). Unfortunately, there is not much awareness among the Institutions and researchers on the parameters of Research excellence in Ranking. The ranking list of Times Higher Education 2020, clearly shows the fall of Indian Institutions’ level of ranking from 200 to 300 range. We have excellence in parameters like teaching and industry income but fall drastically behind in research excellence and citation. While for young institutions, it may be considered an important task to seek the country-wise ranking list (NIRF) and work towards annexation. It will be more meaningful to look into the data chosen for ranking the institutions within the country or region for local decision-making and policy changes. However, in taking any decision, it is important to know for which type of ranking (local, or International) the university is aiming and then develop a policy which enriches their objective. THE (Times Higher Education) and QS (Quacquarelli Symonds) university ranking agencies use not only research metrics like Research, Citation and industry-income but internationalisation and reputation surveys are also as part of their ranking exercise. When the basic objective of an institution is teaching, reputation survey and internationalisation can be achieved easily. The continuous urge to make a better position in the world University ranking influences the researchers, policymakers and administrators of the institutions to develop a strategic atmosphere in the universities. Good quality data is essential to develop ideas and solutions that lead to model development for impact studies, policy planning and performance evaluation. In this paper, we have tried to do some innovative analysis by choosing the valuable metrics of importance. We have selected five institutions, each from the THE top-ranked institutions. The selected institutions are Indian Institute of Technologies, Central Universities, State Universities and Private Universities (Table 1). We have computationally retrieved and analysed the different parameters of research output qualitatively and quantitatively.
2. RELATED WORK
The existing literature shows that there were already many studies performed in connection with the performance evaluation using citation metrics. Reviewing the study on metrics there were many studies done and have been increasingly adopted to assess the scientific impact of papers, authors to journals and institutions5-11. Prathap G. et al. have done extensive study on IITs, private universities and research intensive-higher education institutions in India3, 12-24. Another study done by Singh V. K. et al. was on performance evaluation comparing IITs, and also a comparative analysis on Nanyang Technological university (NTU), Singapore and Massachusetts Institute of Technology (MIT), USA and compared them with Indian Institute of Science (IISc), Bangalore20. The metrics they have used are Total Publication, Citation, Average Citation per paper, Highly-Cited paper and h-index. Discipline-specific research strength was also computed. They have used impact – Citation-Exergy (iCX) trajectory analysis for the evaluation of institutions4,12-24. The study particularly shows that the performance of top-ranked IITs and other institutions are far below than that of some of the top leading institutions in the world, they have chosen for comparison12. In a recent paper, Prathap G. has discussed the importance of authorship contributions in the scientific literature24. The present study emphasises the important bibliometric parameters and have evaluated the collected data using qualitative and quantitative parameters. Institutions first have to set their targeted ranking (National/International) and then needs to formulate strategic planning for achieving the goal. Institutions who are young in the field of ranking has to decide and plan to which ranking they have to participate. The methodology is different for different Ranking agencies and it is advisable to institutions who are young in ranking to start with national rankings. Pursue progress through continuous monitoring and build an appropriate strategy for achieving bigger heights.

3. METHODOLOGY
We have collected the data from the Scopus database from 2008 to 2018 for 20 Indian Institutions. These 20 institutions were considered based on the Times Higher Education World University (THE) rankings 2020. For the benchmarking, we have selected five each of top-ranked (THE) Indian Institute of Technologies (IITs), Central Universities (CU), State Universities (SU), Private Universities (PU) and Research Institutes (RI).

We have collected the data from the Scopus database from

Table 1. The sample taken for the study

<table>
<thead>
<tr>
<th>Indian Institute of Technologies (IITs)</th>
<th>Central Universities</th>
<th>State Universities</th>
<th>Private Universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Institute of Technology, Bombay (IIT-B)</td>
<td>Indian Institute of Science Bangalore (IISc)</td>
<td>Jadavpur University (JU)</td>
<td>Amrita Vishwa Vidyapeetham (Amrita)</td>
</tr>
<tr>
<td>Indian Institute of Technology, Delhi (IIT-D)</td>
<td>Banaras Hindu University (BHU)</td>
<td>Panjab University (Punjab)</td>
<td>Manipal Academy of Higher Education (MAHE)</td>
</tr>
<tr>
<td>Indian Institute of Technology, Kanpur (IIT-K)</td>
<td>University of Delhi (DU)</td>
<td>University of Mysore (MU)</td>
<td>SRM University (SRM)</td>
</tr>
<tr>
<td>Indian Institute of Technology, Kharagpur (IIT-KH)</td>
<td>Aligarh Muslim University (AMU)</td>
<td>University of Pune (Pune)</td>
<td>Thapar University (TU)</td>
</tr>
<tr>
<td>Indian Institute of Technology, Madras (IIT-M)</td>
<td>Pondicherry University (PU)</td>
<td>Annamalai University (AU)</td>
<td>Birla Institute of Technology and Science Pilani (BITS)</td>
</tr>
</tbody>
</table>

2008 to 2018 for 20 Indian Institutions. These 20 institutions were considered based on the Times Higher Education World University (THE) rankings 2020. For the benchmarking, we have selected five each of top-ranked (THE) Indian Institute of Technologies (IITs), Central Universities (CU), State Universities (SU), Private Universities (PU) and Research Institutes (RI).
Universities (SU) and Private Universities (PU) from India with ten years’ data (2008-2018).

IITs are included in the study because they are representing India in the world ranking (THE/QS). IISc is the top-ranked institution from India and the reason for including IISc among the Central University category is because it is an institution funded by the central government. The University of Delhi is considered for the study because DU comes top in THE ranking among the three central universities located in the State of Delhi (University of Delhi, Jawaharlal Nehru University and Jamia Millia Islamia University). Scopus database was used for retrieving the data, and Analytical tool SciVal was used to extract data for the samples. For retrieving the data from Scopus, we have used the affiliation search tab and the search term used is Institutions name. The metrics used for analysis are mixed metrics of quantity and quality parameters. Table 2 explains the different metrics used for analysis.

3.1 Terminologies and Explanations

Few terminologies used for the present study are:

3.1.1 Outputs in Top Citation Percentiles

Outputs in Top Percentiles indicate the extent to which outputs are present in the most-cited percentiles of the data source (SciVal/Scopus). The citation counts that represent the thresholds of the 1 per cent, 5 per cent, 10 per cent and 25 per cent most-cited papers in Scopus per Publication Year are calculated. The Outputs in Top Percentiles metrics are useful to benchmark the contributions towards the most influential, highly cited publications.

3.1.2 Field-Weighted Citation Impact

Field-Weighted Citation Impact takes into account the differences in research behaviour across disciplines. This metric indicates how the number of citations received by an institution’s publication compares with the average number of citations received by all other similar publications indexed in the Scopus database. A Field-Weighted Citation Impact of 1.00 indicates that the publications have been cited at world average for similar publications. A Field-Weighted Citation Impact of greater than 1.00 indicates that the publications have been cited...
more than that would be expected based on the world average for similar publications, A Field-Weighted Citation Impact of less than 1.00 indicates that the publications have been cited less than that would be expected based on the world average for similar publications (SciVal)\textsuperscript{22-23}.

We have computationally retrieved the data, and qualitative-quantitative analysis was done using different parameters of research output to explain the importance of various metrics. The collected data indicates different metrics such as Scholarly output, Cited and Uncited Publications, Field-weighted citation impact, International, National, Institutional, and Single authored collaborative publications and its impacts. To know the impact of publications like minimum output and huge impact, we have computed the uncited publications also. It is one such area which is often overlooked. Even though the total publications of an institution may be huge, the number of cited papers is a matter of quality and visibility. Also, the citation crowd for a paper depends on the selection of journals, selection of proper keywords as well as the visibility of the publication among the peers. To get a better picture of the collaborative pattern of these institutions, we have collected four types of collaborations and studied its impact. The metrics used are International, National, Institutional and Single Authored collaborative publications and its impacts.

3.2 Data Preprocessing

The collected data was analysed using Python programming language. While computationally analysing the data, we found that there are some parameters which are not contributing much effect on our analysis. For this, Dimensionality Reduction Technique, PCA (Principal Component Analysis) is followed to take care of multicollinearity in the data to remove redundant features. For example, we have total publications, total citations and citation per paper, these variables are giving redundant information, and the removal of citation per paper is not going to affect the result of the analysis. Python’s library Pandas is used for data manipulation and analysis. To clean the data and remove the null, unwanted and missing values, we have programatically managed the imported data set using Pandas library. For the better visualisation of the analysed data, Python’s open-source libraries Plotly and Matplotlib are used.

4. RESEARCH OUTPUT AND METRICS

The data used in Table 3 is the total metrics which is considered as quantity-quality parameters. The four sets of institutions are identified for the study based on their performance in Times Higher Education Ranking 2020 and are top-ranked Institutions from IITs, Central, State and Private universities. The parameters chosen are the following: Scholarly Output (SO), Cited Publications (CP), Uncited Publications (UCP), Collaboration International percentage (CIN %), Collaboration National percentage (CN %), Collaboration Institutional percentage (CI %), Collaboration Single Author percentage (CSA %) and its Impact like CIN Impact, CN Impact, CI Impact and CSA Impact.

As quality metrics, we have also computed publications in Journal quartiles by CiteScore percentiles such as Quartile 1 (CSQ1), Quartile 2 (CSQ2), Quartile 3 (CSQ3) and Quartile 4 (CSQ4). The other quality parameters like Outputs in Top Citation Percentiles C1 per cent, C5 per cent, C10 per cent and C25 per cent are also considered. Field-Weighted Citation Impact is computed for the evaluation of quality. It is a normalised metrics and normalised as 1 for the world average.

4.1 Scholarly Output, Cited and Uncited

We have considered different metrics for analysis, and the first metric is total scholarly output for ten years and cited and uncited documents. The sample institutions are analysed plotted in Fig. 1, to get a clear picture of total publication number, cited and uncited publications. The data plotted using python graphics library Plotly. The highest number of publication comes from Central Universities (CU), followed by IITs and Private Universities (PU).

The contribution of State Universities (SU) is less compared to others, and the highest publication number comes from IITs. However, when we explored the cited and uncited publication count, we could see that 83.34 per cent of Panjab University publication are cited, and it clearly shows the quality of work produced by a state university. This state university can be considered a model for all other state universities to develop a better strategy for the coming years. IITs and few Central University are also coming in the range 83 per cent to 79 per cent cited publications. Private Universities are contributing more on the publication numbers but are counted high in the category of uncited publications (39 % - 22 %) than cited publications in comparison with others. It clearly shows that private Universities need to formulate a better
Figure 2. Publication output growth and predictions for four sets of Institutions.

Figure 3. Differences in collaborative patterns: *CIN, CN, CI, CSA.
strategy to handle this situation. Central universities, IITs and Private universities are contributing more on quantity for the last many years. When analysing the quality parameter, IITs and Central Universities showing consistency in acquiring citations throughout the last five years. Apart from Panjab University, other state universities need to revisit their strategy in the contribution of quantity as well as quality.

4.2 Research Productivity and Five Years’ Growth trend
The Total Publication (TP) graph for the years (2008-2018) is plotted along with the publication trend for the coming five years in Fig. 2. For this statistical analysis, the linear trend model is used, for five years of future trend prediction, the data from the year 2008 to 2018 was considered. The coefficient of determination for regression line accuracy (R²) value is high in all the three sets of institutions and is less in the case of State Universities (SU). In this analysis, Private Universities contribution to Scholarly Output is exceptional in comparison with others.

In 2008, Private Universities contribution was lower to State universities, but we can see an outstanding growth from 2015 onwards, and the trend analysis is showing an increasing tendency for the next five years. SU is not showing much improvement from 2008 to 2018. The predictive analysis is based on the performance of the Universities in the last ten years and State Universities had to work more to reach to the top.

4.3 Collaboration and Its Impact
When analysing the quantity-quality parameters of the four sets of institutions considered for the study, we have observed the different collaborative pattern and its impact. We have selected four types of collaborative publication patterns which are International(CIN), National(CN), Institutional(CI) and Single authored publications (CSA).

We have retrieved the four collaborative publication sets for the four sets of institutions to see the number of publications distributed among all types of collaborations. Then we have analysed the International Collaboration Impact (CIN Impact), National Collaboration Impact (CN Impact), Institutional Collaboration Impact (CI Impact) and Single-Authored Collaboration Impact (CSA Impact). The percentage of International Collaborative (CIN) papers are more with Panjab University (29.8 %) than other universities, and the impact also finds higher (36.7 %) than others. National Collaborative (CN) publications are more with Jadavpur University (48.7 %) and Pune University (47.8 %) than other selected institutions, but the impact is more for IISc Bangalore and IITs. Among the Private Universities, CIN number is less for Amrita (12.3 %) but the impact is very high (25.9). This metric affects the overall Field-weighted citation Impact of an Institution. Compared to the other types of collaborations, publications coming from the CI are more, but the impact is high for CIN and CN and very low for CI. Single authored publications are less in the case of all Institutions except Delhi University (10.2 %), but the impact is very high in the case of Thapar university (16.3 %). This analysis clearly shows that for achieving greater impact and high Field-Weighted Citation (FWCI), Indian Universities must concentrate more on CIN and CN. The international outlook of Universities must be promoted by collaborating more with foreign universities and the Indian government should have a policy for the promotion of this. Moreover, a strategic plan should be formulated to get more impact on the publications with National Collaboration.

4.4 Quality Parameter - Journal Quartile Publication
The publications in journal quartiles are very important in
Figure 5. Output in Top Citation Percentile.

The trend of journal selection in Indian Institutions shows that the major publications are going for Quartile1 and Quartile2 journals. The selection of a journal for publication in quartile1 shows the intensity of research culture in an institution and among the researchers. The awareness of journal quality and selection of suitable journal in the subject area of research is not much in Indian Universities. It is essential to have research enrichment classes to early carrier researchers to build the research culture in the campus. Selecting the Quartile1 to Quartile4 journals for publication shows the quality of the publication and most of the time, researchers who are young in the field of research will not be aware of the quality of journals. It is necessary to select journals with quality to explain the importance of the work done in the paper.

Figure 4 shows the quantity of publication distribution in each quartile. Journal quartiles are calculated using citations and publications for the previous three years. IISc, Bangalore has the highest publication number in quartile1 journals. The visibility of any publication lies in the selection of journals in suitable subject areas. The selection of journals in the top quartile in the appropriate subject area will lead to more impact. That is the reason for the contribution of most of the publications of Panjab University in Quartile 3 shows better impact.

4.5 Output in Top Citation Percentile

A very important quality parameter is Output in Top Citation Percentile. This evaluation clearly shows the quality of the publication. The data plotted in Fig. 5, give a clear picture of the percentage of citations in each quartile. Outputs in Top Citation Percentiles indicate the extent to which outputs are present in the most-cited percentiles of the data source. The citation count that represent the thresholds of the 1 per cent, 5 per cent, 10 per cent and 25 per cent most-cited papers in Scopus per Publication Year are calculated. The time period referred here is the year in which the papers were published not the year in which citations were received. The above metrics are useful to benchmark the contribution towards the most influential, highly cited publications of an institution.

Figure 5 is self-explanatory, and it explains that the University of Panjab has the most cited percentile in all followed by Thapar University, Delhi University, Aligarh Muslim University, IIS Bangalore etc. Any institution needs to fall in the top1 per cent citation in the published year to make a permanent mark in any world university ranking.

4.6 Field-Weighted Citation Impact

Another indicator to assess the University’s citation performance is the Field-Weighted Citation Impact (FWCI).

Figure 6 shows Panjab University’s FWCI is the highest (1.54) among other Institutions. Another indicator to assess the University’s citation performance is the Field-Weighted Citation Impact (FWCI). As seen in Fig. 6, Panjab University’s FWCI is the highest (1.54) among other Institutions. The observation is that the citation impact of the publications from Panjab University, in their field, is more than the world average. FWCI is a normalised metric, and it is normalised to 1 as the world average. Twelve well-known institutions’ citation impact is more than the world average in our analysed data set. However, other institutions have to work on parameters which enhances their FWCI.

5. CONCLUSIONS

In summary, the combined metrics explain the factors which has to be taken into account seriously while approaching the world University ranking. It is not only the huge number of Scholarly Output that matters but the Cited and Uncited...
publications also matter when computing Quality upon Quantity. In terms of Quantity, the statistical growth prediction in the linear trend graph in Figure 2 has predicted that future growth is promising. Nevertheless, it is necessary to reduce the Uncited publication number from the denominator to improve the Citation per Publication.

Our main finding in the study is the impact of Collaborative publications and the impact of Citation in Quartiles. Almost all the Top-Ranked institutions taken for the present study have the publication number more from Institutional collaboration than National or International Collaborations. This metric affects the overall Field-Weighted Citation Impact of an intuition. The Impact study of these collaborations clearly shows that International Collaboration impact is higher when compared to the National Collaboration impact. To improve National Collaboration Impact strategic plan should be formulated, and efforts must be more on improving the impact. Institutional Collaborative papers must reduce and international, and national level collaboration must increase with impact. The visibility after publication is more important than restricting the expertise within our purview by increasing institutional collaborations.

The shortcoming in the study is that the subject wise/field-based publication number and impact is not analysed. It was assumed that the present study is based on the world University ranking and overall research, citation, teaching and other parameters. And the more important aspect which is having weightage is research and Citation. IITs are highly specialised in terms of discipline, and their research is focused on specialised topics. The State, Central and Private universities are being multi-subject disciplined institutions show the tendency of fluctuation in selecting research areas and publishing research. It can be a future study by taking into account the subject area-wise data.

In conclusion, our study has identified important metrics which can be considered for developing a strategic plan and policy to achieve high performance in research. It is a fused index metrics which explains along with Scholarly Output; Uncited Publication numbers have to be reduced drastically to achieve better impact. With Total Citation parameter, it is important to check which Collaboration gives more Impact to the Institution and focus on such collaboration, avoiding other collaboration patterns which does not benefit the progress of the institution much. If we consider publication in Journal Quartile, Output in Top-Cited Percentile and factors affecting the FWCI, a combination of these metrics can contribute to a strategic approach to build a new research policy for your institution to go for ranking.

ACKNOWLEDGEMENT
The corresponding author is thankful for the financial assistance from the DST project entitled “Influence of Government Policies and Funding on the Research Output of Government and Private Institutions in INDIA: A Bibliometric and Scientometric Study, Ref. (No.DST/NSTMIS/ 05/41/2017-18 dated 28.02.2019)

For data and metrics, we are thankful to Scopus and SciVal.

REFERENCE
24. Prathap, G. Rationalisation of the SS Bhatnagar Prizes scheme (Correspondence), 2014.

CONTRIBUTORS

Ms Sheeba Pakkan is a Bibliometrician at Manipal Academy of Higher Education, Manipal. Her area of interest is to provide high-level expertise, and advice to the University on the use of bibliometrics and related indicators, University Ranking, Data analysis, and Research Metrics.

Her individual contribution in the current study: Conceptualization, data retrieval and overall contribution to the present study.

Dr Christopher Sudhakar is a Director of Quality at MAHE, Professor MCON, MAHE. He has 34 years’ diverse experience in teaching and learning, clinical, academic administration, infection control, research and quality assurance in India and abroad.

His individual contribution in the current study: Guide, editor and advisor to the present study.

Mr Shubham Tripathi is a Data Analyst who is working at Manipal Academy of Higher Education in the DST. He has done his Masters of Engineering in Big Data and Data Analytics with the ability to deliver valuable insights via data analytics and advanced data driven methods.

His individual contribution in the current study: Data Preprocessing, Data Analysis and Visualization for the current study.

Dr Mahabaleshwara Rao is currently working as Associate Professor and Deputy Chief Librarian at Dept. of Library & Information Science and Health Sciences Library, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka. He holds his Ph.D. degree in library and information science from Mangalore University, Mangalore specialization being on user study. His areas of interest include Health Sciences Librarianship, Web Resource and Information Literacy.

His individual contribution in the current study: Guide and advisor to the present study.