ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/354347648

Fine-grained academic rankings: mapping affiliation of the influential researchers with the top ranked HEIs

Article *in* Scientometrics · September 2021

DOI: 10.1007/s11192-021-04138-z

CITATIONS
0

READS
50

2 authors:

Muhammad Sajid Qureshi
Foundation University
3 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Ali Daud
University of Jeddah
101 PUBLICATIONS **1,554** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Academics and Research View project

Project Scientometrics View project

All content following this page was uploaded by Ali Daud on 07 September 2021.

The user has requested enhancement of the downloaded file.

Scientometrics
https://doi.org/10.1007/s11192-021-04138-z

Fine-grained academic rankings: mapping affiliation of the influential researchers with the top ranked HEIs

Muhammad Sajid Qureshi[1,2] · **Ali Daud**[1,3]

Received: 3 December 2020 / Accepted: 18 August 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract

The academic ranking process has considerably evolved in the past fifteen years and the evolution has gained the momentum in last few years. Starting with the holistic rankings of world universities in 2003, it has crossed the milestone of subject-specific rankings. Nevertheless, the academic rankings published by even the reputed ranking entities are facing various criticism, in terms of their *transparency, validity,* and *coverage*. This research effort focuses on enhancing the credibility of the ranking process through the fine-grained analysis of the academic data. The proposed fine-grained analysis drives the researcher's profiles from the Google Scholar Citations repository. While the DBpedia repository is employed for the information about HEIs and countries. The influential researchers are identified using the *ResRank* methodology. While, for consistent comparison of the subject-specific rankings of global HEIs, the *Grand Average Rank* (GAR) metric is employed. The resultant academic rankings with respect to the *Research Faculty, Research Productivity,* and *Research Impact* make the ranking process more transparent and fine-grained. The analysis also helps in understanding the causes of differences among the academic rankings published by the ARWU, THE, and QS rankings systems. The growing interest in the subject-specific and sub-discipline-specific rankings is irreversible. The fine-grained analysis is a response to the need.

Keywords Academic rankings · Subject specific rankings · Bibliometrics · Google scholar citations · Linked data · DBpedia

✉ Ali Daud
alimsdb@gmail.com

[1] Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan

[2] Department of Software Engineering, Foundation University Islamabad, Islamabad, Pakistan

[3] Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

Published online: 02 September 2021

Introduction

The role of the Higher Education Institutions (HEIs) as a catalyst of the national and global development process, has become established worldwide. Governments of the developed countries nurture research, innovation, and commercialization through the institutions. Moreover, they also have adopted systematic ways to gauge and evaluate the academic performance of their HEIs. Usually, the evaluations declare the results in terms of the ranking scores. Various public and private ranking systems are gauging and publishing the HEI's rankings, including the Academic Rankings of World Universities (ARWU) from China, Quacquarelli Symonds (QS) from the United Kingdom, Times Higher Education (THE) from the United States, CWTS Leiden Rankings from the Netherlands, and University Rankings by Academic Performance (URAP) from Turkey.

These ranking systems have been publishing the rankings results since 2003 and their outcome has gained considerable acceptance by the major stakeholders (Federkeil, 2011; Qureshi et al., 2021). The interest of academic experts in the rankings process also has increased in the last ten years. Bibliographic data repositories are reporting a significant increase, in the publication of research articles focused on the academic ranking process. According to a study, there were only five international university ranking systems before 2010. Nevertheless, today the number has increased to seventeen. Similarly, in the year 2009, the academic experts published fewer than twenty journal articles on the topic; while in the year 2019, they published more than 100 articles, according to the Scopus database (Lokman, 2020). The increasing interest in the academic rankings is understandable because of its various applications as discussed in Qureshi et al. (2021); Abramo et al., 2016). Overall, academic ranking has become an inevitable part of higher education, nationally and internationally.

HEIs' ranking systems significantly have improved their quality during the past fifteen years. Currently, they are more decision supportive and user-friendly than they were a decade ago. Nevertheless, the ranking results need to be more transparent and fine-grained, as highlighted in Qureshi et al. (2021; Fredrik & Gunnar, 2016; Lin et al., 2013; Jo, 2019). The related literature highlights various controversial issues of the reputed rankings. Major of these issues are controversial ranking methodologies, validity and transparency of the ranking results (Aguillo & Bar-Ilan, 2010; Friso et al., 2020; Henk & Moed, 2016), the use of the "subjective" ranking indicators, misinterpretation of the academic performance (Fredrik & Gunnar, 2016), and limited coverage of the global HEIs (Garcia & Herrera 2013).

We examined the holistic ranking scores allocated to the top-performing HE, Is in the years 2017, 2018 and, 2019 by the three reputed ranking systems—ARWU, THE, and QS. To obtain a more stable ranking score for each of the HEI, we also computed the 3-year average of the ranking scores, for each of the systems. Furthermore, for descriptive analysis, we also computed the average of the nine ranking lists (Grand Average Rank–GAR Score), produced by the three ranking systems in the stipulated years. A summary of the analysis is presented in Table 1 and a part of the results is visualized in Figs. 2 and 4 (Table 2).

For instance, the USA-based Stanford University (SU) stands first according to the GAR score. In other words—unanimously ranked at the first position, by ARWU, THE, and QS in the years 2017, 2018, and 2019. Nevertheless, the university attains significantly different ranks with respect to (w.r.t.) the various ranking dimensions. For example, as shown in Fig. 1, its rank w.r.t. the Influential Research Faculty (IRF) in the Engineering discipline is 2nd, while it has the 16th, 17th and 18th positions in Economics and Management, Health

Table 1 HEIs' Average Ranks in 3-years by the three ranking systems and their GAR scores

	Higher Education Institution (HEIs)	Acronym	ARWU			THE			QS			3-Year Rank Average			GAR Score
			2017	2018	2019	2017	2018	2019	2017	2018	2019	ARWU	THE	QS	
1	Stanford University	SU	2	2	2	3	3	3	3	2	2	2.00	3.00	2.33	2.44
2	Harvard University	HU	1	1	1	6	6	6	2	3	3	1.00	6.00	2.67	3.22
3	Massachusetts Institute of Technology	MIT	4	4	4	5	5	4	1	1	1	4.00	4.67	1.00	3.22
4	University of Cambridge	UoC	3	3	3	4	2	2	4	6	5	3.00	2.67	5.00	3.56
5	University of Oxford	UoX	7	5	6	1	1	1	6	5	6	6.00	1.00	5.67	4.22
6	California Institute of Technology	CIT	9	9	9	1	3	5	5	4	4	9.00	3.00	4.33	5.44
7	University of Chicago	UCh	10	10	10	10	9	10	4	6	5	10.00	9.67	5.00	8.22
8	Princeton University	PrU	6	6	6	7	7	7	11	13	13	6.00	7.00	12.33	8.44
9	University of Pennsylvania	UoP	17	19	19	13	10	12	6	5	6	18.33	11.67	5.67	11.89
10	Yale University	YU	11	12	11	12	12	8	15	16	15	11.33	10.67	15.33	12.44
11	Swiss Federal Institute of Technology (ETH) Zurich	ETH	19	19	19	9	10	11	8	10	10	19.00	10.00	9.33	12.78
12	University College London	UCL	16	17	15	15	16	14	7	10	7	16.00	15.00	8.00	13.00
13	Columbia University	ClU	8	8	8	14	16	14	20	16	18	8.00	14.67	18.00	13.56
14	Imperial College of Science Technology and Medicine London	ICL	27	24	23	8	8	9	9	8	8	24.67	8.33	8.33	13.78
15	Cornell University	CrU	14	12	13	19	19	19	16	14	14	13.00	19.00	14.67	15.56
16	University of California at Berkeley	UCB	5	5	5	10	18	15	28	27	27	5.00	14.33	27.33	15.56
17	Johns Hopkins University	JHU	18	18	16	17	13	12	17	21	17	17.33	14.00	18.33	16.56
18	The University of California at Los Angeles	UCLA	13	11	11	14	15	17	31	33	32	11.67	15.33	32.00	19.67
19	University of Michigan	UMi	24	20	21	21	21	20	23	21	20	21.67	20.67	21.33	21.22
20	Duke University	DkU	26	26	28	18	17	18	24	26	21	26.67	17.67	23.67	22.67

Table 2 Academic ranking dimensions in ARWU, THE, and QS ranking methodologies

Ranking Dimension	THE	QS	ARWU	UniRank
Research Productivity	65	20	40	45
Academic Quality	30	20	50	35
Academic-Sustainability	–	–	–	15
Academic Peer Review	–	40	–	–
Employer Reputation	–	10	–	–
Per Capita Performance	–	–	10	–
International-Factor	5	10	–	5
Total	100	100	100	100

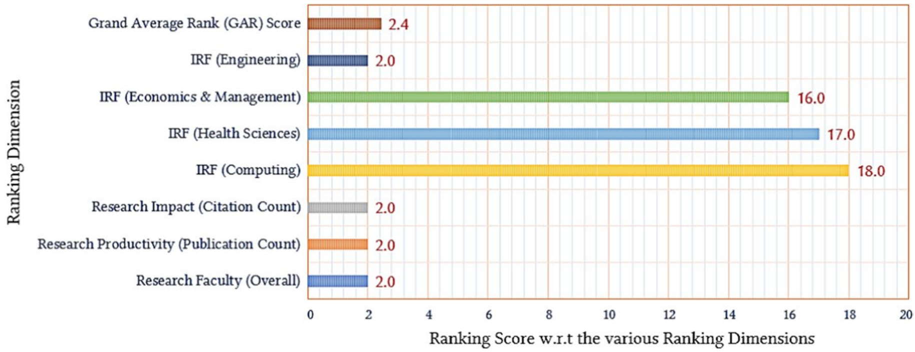

Fig. 1 Stanford University positions w.r.t. various ranking dimensions

Sciences, and Computing, respectively. Nevertheless, the rank of the university w.r.t. the *research productivity,* and *research impact,* and *research faculty* are consistent with the GAR score.

Nevertheless, HEIs with a higher GAR score, have significantly different ranks w.r.t. the various ranking dimensions. Such discrepancies shake the confidence of the end-users in the ranking results and provide supportive arguments to the critics. The users also feel a need to have a more fine-grained analysis of the HEI's performance evaluation. As compared to the holistic ranking score, the fine-grained analysis of the academic ranking data provides more representative rankings of the global HEIs. Such analysis would prove more decision-supportive due to the expansion in the limited picture of the global HEIs. For more detail, please consult Table 1. This analysis also gave multiple insights into the subject-specific rankings of HEIs in the four scientific disciplines.

Research objectives

Use of the substantially different ranking methodologies, and the variety of data sources having considerable inconsistencies, are the main causes of the inconsistencies in the outcome of the reputed ranking systems. The subjective ranking indicators (Huang et al., 2020) also help in the manipulation of the ranking results. These differences are among the open issues of the rankings and need to be addressed. Nevertheless, finding suitable solutions for all the issues requires extensive time, effort, and resources. We already have done a research effort focusing credibility of the ranking process. There we proposed an academic ranking methodology based on objective ranking indicators sourced from transparent data sources. This work was reported in Qureshi et al. (2021). The *OpenRank* methodology suggested in the research was developed for the holistic ranking of the HEIs. Nevertheless, during the research effort, we seriously felt the need for the fine-grained analysis at the subject-specific level.

Research contributions

This research work is focused on two of the issues—*validity,* and *transparency* of the ranking process, with an in-depth analysis of the academic rankings in the four scientific disciplines. Significant contributions of the research work, are delineated as follows:

- A fine-grained analysis of the academic data w.r.t. the *research faculty*, *research productivity*, and *research-impact* of the top-performing HEIs, using publicly verifiable data sources.
- Producing the subject-specific rankings of the global HEIs w.r.t. the influential research faculty, by mapping the affiliation of the influential researchers with the top-performing HEIs, to explore their research productivity and research impact.
- Demonstrating the effectiveness of the fine-grained data analysis strategy, using publicly verifiable data repositories i.e., DBpedia[1] and Google Scholar's Citations.[2]
- Considering the context of global HEIs in the academic ranking process to obtain a more realistic ranking.

In this article, the rest of the content is organized into sections. The significant findings during the literature review are reported in section "Related work. The research problem is formulated in section "The Research Problem". while section "Methodology" presents the proposed ranking methodology. The experiments and their results are reported in section Experiments. Finally, Section "Conclusions and future work" concludes the research work.

Related work

Keeping in mind the research objectives—fine-grained analysis of the ranking parameters and use of the publicly verifiable data sources, we focused the research efforts of two types. First, that was aimed at the fine-grained analysis of the academic ranking dimensions. Second, the academic rankings efforts that employed publicly verifiable data sources. An abridged review of the research efforts is following.

José Luis analyzed the relationship between research impact and the organizational structure of co-author networks using the evolving bibliographic data source, Microsoft Academics (Ortega, 2014). According to the analysis, the dedicated researchers who are part of the sparse and small research networks, have higher research productivity and

[1] http://dbpedia.org/page/DBpedia/.

[2] https://scholar.google.coms.

better research impact. In other words, the HEIs with better facilitation for the dedicated researchers, attain high research-productivity and high impact research.

Bo (Yang et al., 2015), explored the relationship between the leading scientists and top-performing organizations. They argued that up to 80% of the world's highly influential researchers work at the top-performing organizations, especially in large fields such as materials science, physics, chemistry, neuroscience, and health sciences. In general, top-performing institutions have the competitive advantage of having excellent researchers; only a few exceptions diverge the trend.

Giovanni Giovanni & Ciriaco (2015) argued that ranking organizations or countries based on either the total number of Highly Cited Articles (HCAs) or by the ratio of HCAs to total publications, is not a substantial way to assess their productivity. The authors proposed a single influential indicator *HCA Per Scientist*, to gauge the research productivity of HEIs. They applied the bibliometric measure HCA, in finding research excellence of Italian universities in each field and discipline of the hard sciences. According to them, the indicator is more effective, timesaving, and less costly.

The research work reported in Qureshi et al. (2021) by M. S. Qureshi et. al. emphasized the need of employing objective and publicly verifiable data sources in the academic ranking process. The researchers highlighted a few of the causes of the differences among the ranking scores published by the reputed rankings entities, for similar global HEIs in the years 2017–2019. According to their findings, the use of objective ranking parameters and transparent data sources would result in more consistent rankings and thereby enhance the credibility of the academic ranking process. The authors also proposed a new academic ranking methodology—OpenRank to demonstrate the feasibility of the objective and publicly verifiable data sources. The intended fined-grained analysis and the sub-discipline specific rankings of HEIs is an extension of the research effort.

John Mingers et al. examined 130 HEIs of the UK by using Google Scholar Citations (GSC) data. They used citation-based statistics to produce the HEIs' rankings (Mingers et al., 2017). For the same HEIs, the authors made a comparison between resultant rankings and those produced by the UK Research Excellence Framework.

The authors claimed more credible results as compared to the ranking results by REF, with additional benefits of cost-effectiveness and efficiency. The ranking-indicator-oriented analysis provided better insight into the performance evaluation.

The analysis presented in Jabnoun (2015) by Naceur Jabnoun, explored the influence of wealth, transparency, and democracy on the number of universities per million people ranked among the top 300 and 500. the analysis revealed that countries with top-ranked universities had higher Gross Domestic Product (GDP) Per Capita, better transparency and democracy levels than countries with no top-ranked universities. The author highlighted the fact that university management, like the management of any other organization, is influenced by environmental factors including political and economic factors. The analysis emphasized the need of considering the context while ranking and comparing global HEIs. Such context sensitive analysis is another supportive voice for the proposed fine-grained analysis strategy.

The work presented in Sheeja et al. (2018) by Sheeja et al. highlighted role of the *Research Productivity* in the academic ranking process. They studied the case of Indian HEIs which are highly ranked as per the National Institutional Ranking Framework (NIRF) of India. According to their analysis, the most decisive performance indicator of an HEI is the research productivity. The authors employed the quantitative parameters after extracting them from the official websites of NIRF, THE, and QS.

Jun et al. (2018) claimed improvement in predicting the impact of research publications in the heterogeneous temporal academic network. Their proposed method *Personalized Prediction of Scholars' Scientific Impact* (PePSI) classifies the researchers into different types according to their citation dynamics. It then predicts their impact in heterogeneous temporal academic networks, by applying different fit-functions to represent their citation-dynamics that vary with time. PePSI claimed the best performance in the identification of outstanding researchers in the shortest time. Such research work enables the fine-grained analysis of the scholar's research impact.

Koen Frenken et. al. empirically analyzed university research performance in terms of *research excellence*, *internationalization*, and *innovation* (Koen et al., 2017). The work highlighted that these indicators are prone to conceptual ambiguity and uncertainty; nevertheless, in many cases, students and, HEI's manager consider them increasingly meaningful. Moreover, the difference in size, disciplinary orientation, and country location are major causes of the difference among the HEIs' research performance. This suggests that instead of simple global benchmarking, a more fine-grained benchmarking is meaningful among HEIs having similar characteristics. The performance evaluation should consider the contextual information of an HEI, its mission, orientation, and geographical location.

Baris Uslu in Uslu (2020) used the predictive approach of correlational research to drill downed methodologies of the four academic ranking systems –ARWU, THE, QS and URAP. He employed the ranking results published by these systems in year 2018, to specify an expanded set of the prominent ranking indicators. In his opinion, these potential indicators cover the common aspects of the rankings systems and would help the university leaders to develop better strategies. The researcher also determined the percentage values of the indicators including the *Citation, Income, Internationalization, Prize, Publication, Reputation,* and *Ratios/Degrees.* According to findings of the analysis *Research Reputation* contribute 73.71% to universities' ranking scores. While, amazingly, the *Income* is the only negative contributor with a weight of- 1.78%. The research also revealed that in the HEIs ranking based on the new weights, only 19 universities occupy the same position among the 224.

universities. This research work also recommends use of the objective ranking parameters; therefore, it consolidates the philosophy behind our fine-grained analysis strategy.

Overall, the academic ranking efforts, covered in the literature review, emphasize the need for a more fine-grained analysis objective and context-sensitive performance evaluation of the of the HEIs. Such analysis would better help the various academic stakeholders in their decision-making process. The academic managers would focus area of their interest, considering the analysis.

Problem framework

This section briefly describes some terminologies and introduces the three reputed ranking systems in brief. Readers having the background knowledge can skip the preliminaries and introduction of the systems. Nevertheless, the methodology requires due attention for a better understanding of the analysis.

Fundamental terminologies

The process of academic rankings usually involves a complex ranking methodology. The methodology consists of the ranking indicators that reflect the objectives of the ranking entity. The indicators differentiate a ranking system from others. One or more indicators gauge or measure effectiveness of a particular ranking dimension. The indicators may be *Objective* or *Subjective*. For example, the "Faculty/Student Ratio" is an objective indicator. Whereas the "Employer Reputation" is an example of the *Subjective* parameters (Maxime & Alex, 2018b). An effective academic ranking methodology assists various stake holders in the decision-making process in academia. The methodology also helps the academic managers in the better management of the areas of interest for the HEI. It also provides decision-supporting analytics to the sponsors of the HEIs (Boulton, 2011). An effective academic ranking prefers quantitative indicators over qualitative indicators. It provides a fine-grained analysis based on the transparent data sources, to the end users. Scopus, DBpedia, GSC, and ArnetMiner repositories are examples of publicly verifiable data sources.

Some well-known heis ranking systems

Multiple academic rankings systems have been publishing the rankings of HEIs worldwide for the last fifteen years. However, we focused on the three widely observed academic ranking systems—ARWU,[3] THE,[4] and QS.[5] These systems are well-established and have been publishing the global HEIs rankings for more than ten years. Despite having differences, their ranking methodologies are mature and well-defined. These systems are discussed in detail in Qureshi et al. (2021); Fredrik & Gunnar, 2016; Friso et al., 2020). Therefore, we shorten their introduction. The subject-specific ranking methodologies of the three systems can be consulted on their official websites through the URLs provided in the footnotes.

The research problem

According to the literature (Fredrik & Gunnar, 2016; Jo, 2019; Lin et al., 2013; Qureshi et al., 2021), academic experts have highlighted various issues of the ranking process. The prominent of them are listed below:

(a) Controversial ranking criteria or validity of the HEIs' rankings
(b) Misinterpretation of the academic performance
(c) Unavailability of the data
(d) Use of data sources that are not transparent.
(e) Limited coverage of the global HEIs
(f) Use of subjective ranking indicators in the ranking methodologies.

[3] http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/Methodology-for-ShanghaiRanking-Global-Ranking-of-Academic-Subjects-2020.html.

[4] https://www.timeshighereducation.com/world-university-rankings-2021-subject-computer-science-methodology.

[5] https://www.topuniversities.com/subject-rankings/methodology.

Addressing all the above-mentioned issues is beyond scope of this research work. This research work maps the affiliation of the influential researchers with the top-performing HEIs, to explore their research productivity and research impact. It also demonstrates the effectiveness of the proposed data analysis strategy using publicly verifiable data sources i.e., DBpedia and Google Scholar's Citations repository.

Methodology

In this section, we propose the methodology to gauge the impact of influential researchers on the rankings of HEIs. Next, we discuss the major components employed in the methodology. The fine-grained analysis strategy involves three phases. First, the selection of noteworthy ranking indicators for extracting influential researchers from the specified data sources. Second, extraction of the profiles of influential researchers from the data for the selected ranking indicators. Finally, mapping affiliations of the influential researchers with the HEIs that have been continuously ranked among the top-100 s in the three stipulated years, by the three well-known rankings systems including the ARWU, THE, and the QS.

Data sources

A sound and widely applicable ranking methodology employ the objective and publicly verifiable data sources. These quality checks on data make various existing academic ranking methodologies controversial (Garcia & Herrera 2013; Fredrik & Gunnar, 2016). In the proposed fine-grained analysis, we employed multiple data sources that satisfy the necessary conditions. A brief description of them is following.

Official ranking results by the reputed rankings systems

As per the research objective, for fine-grained analysis of the academic ranking statistics, we reused the existing ranking results by the three well-known ranking systems—ARWU,[6] THE,[7] and QS.[8] These ranking results are publicly available on their official websites. Some of the systems provide a limited amount of data in spreadsheets or flat files.

Google scholar citation database

Google Scholar Citation (GSC) repository was created and is currently owned by Google incorporation. It was launched on November 20, 2004.[9] GSC allows its users to search for academic resources in various publicly accessible digital libraries. Peer-reviewed online academic journals, digital books, conference papers, research theses, technical reports, and other scholarly literature are covered by the GSC index. Research publications published by research scholars are the most significant part of the repository (Luis, 2014).

6 http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/index.html.
7 https://www.timeshighereducation.com/world-university-rankings/2021/world-ranking.
8 https://www.topuniversities.com/subject-rankings/2020.
9 https://en.wikipedia.org/wiki/google_scholar.

Table 3 Researcher's Information extracted from the GSC repository

Indicator	Acronym	Description
Research Productivity	PC	Author's research Productivity gauges the research publicity
Research Impact	CC	Representation of the author's research impact
H-Index	HI	Value of the H-Index of an author
Authors Per Paper	APP	Author's Research Contribution in his publications w.r.t. count of the co-authors
Author(s)	Auth	Name(s) of the authors of a research publication
Author's Affiliation	Affiliation	Name of the research institution, with which the author is currently affiliated
Research Interests	RI	Name(s) of the research areas representing the research interests of the author

In our research work, we extracted various statistics about the influential researchers from GSC. Some of the data attributes are *Author-Names* (Auth), name of the current affiliation of the author (Affiliation), *H-Index* (HI) score, *Publication Count* (PC), *Citation Count* (CC), and *Author Per Paper* (APP) Score. It is worth mentioning that by default, the GSC orders researchers' profiles according to their research impact or citation count. A brief but comprehensive description of the information extracted from GSC is presented in Table 3. GSC maintains profiles of hundreds of thousands of researchers, moreover, the count is increasing every day. Moreover, the GSC repository also has data quality issues. Therefore, adopting a careful data reduction approach, we selected only the top two hundred researchers' profiles w.r.t. the *Citation Count* (or research impact). To ensure substantial *Publication Count* (or research productivity), we selected the profiles having at least 100 publications. For a more generalized analysis, we selected the profiles from the four scientific disciplines—Computer Sciences (CS), Health Sciences (HS), Engineering Technology (ET), and Economics and Management Sciences (EMS). These four disciplines are substantially common among the study programs offered by the global HEIs. Moreover, the four disciplines have considerable data representation in the GSC repository. In the future, we would consider other scientific disciplines as well. The intended data analysis is based on the GSC data snapshot taken in November 2019.

The DBpedia dataset

The DBpedia repository is maintained by Wikipedia. It provides a complementary service to Wikipedia by compiling knowledge from more than a hundred projects of Wikimedia. DBpedia is crowd-sourced by thousands of contributors. The DBpedia project was officially launched in 2007 at the World Wide Web conference. DBpedia is among the significant data repositories in the Linked Open Data cloud (Heath & Lee, 2009; Meymandpour & Davis, 2016). The Linked Data cloud is constituted by hundreds of global digital repositories on the Internet. The cloud contains structured data related to various domains like Geography, Healthcare, Media, and Academia (Meymandpour & Davis, 2016). The objects or entities in the cloud are described semantically and structured according to the Resource Description Framework (RDF).

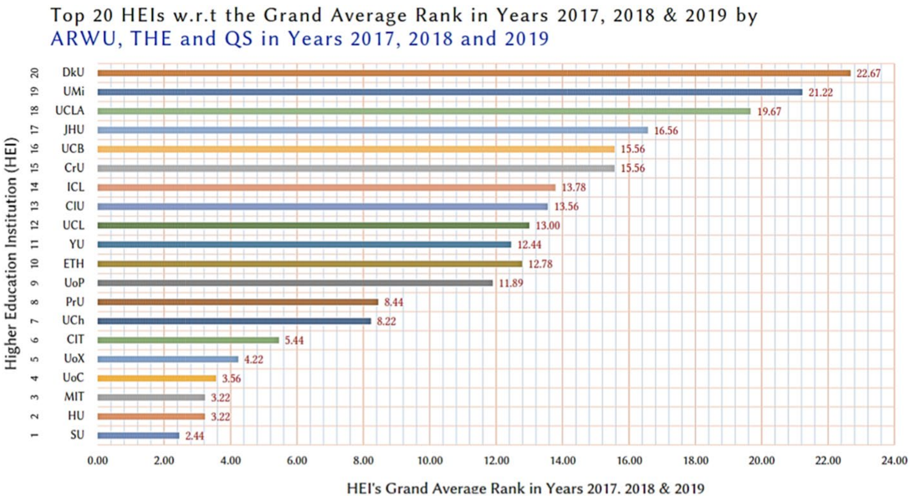

Fig. 2 Top 20 HEIs w.r.t. the GAR Score, in Years 2017–2019

The English version[10] of the DBpedia currently describes millions of "things" and their affiliated "facts". These things are classified with a consistent ontology. The information can serve many existing data-hungry and innovative data products. The DBpedia contains meaningful information about the HEIs, researchers, research publications. In our research work, we employed the DBpedia repository by posing the SPARQL (Heath & Lee, 2009; Holst & Höfig, 2013) queries to a SPARQL-Endpoint to extract information related to the HEIs and countries. A sample of the SPARQL queries for DBpedia, is uploaded to the relevant repositories at GitHub.[11]

The ranking measures

Heading towards the fine-grained analysis of the academic performance evaluation data, we adopted a systematic approach. Initially, we generated a list of the HEIs with top performance in the three years 2017, 2018, and 2019 using the ranking results published by the three academic ranking systems—ARWU, THE, and QS. As a part of the data reduction, only the top hundred HEIs were selected from each of the ranking lists. Then, for each of the ranking systems, we determined the HEIs that repeatedly maintained their position in the list of top-100, in the three consecutive years.

This list was an intersection of the three rankings published by a system. This process was repeated for each of the three ranking systems, to obtain their top-100 HEIs. After finding the three lists of top-100 HEIs, we simply took the union of the three lists. This gave us the final list of top-performing HEIs in the stipulated three years. To obtain a stable ranking score for an HEI, we computed the 3-year average of the raking score, for each of the three ranking systems.

10 https://wiki.dbpedia.org/services-resources/datasets/data-set-39.

11 https://github.com/muhammadsajidqureshi82/OpenRank.git.

Comparison of the 3-year average scores also highlighted the significant differences among the rankings systems even for a single HEI. We also calculated the Grand Average Rank (GAR score)—the overall average of the nine ranking lists for each of the HEIs, based on the rankings published by the three systems, in the three years. The GAR score is a rigorous criterion for identifying the top-performing HEIs.

Although, each of the three ranking systems, publishes a list of the top 100 global HEIs; Nevertheless, the analysis showed that only 40–50 HEIs are commonly ranked as the top-performing HEIs by the three systems. HEIs having a ranking score of more than 50, take considerably different positions in the three rankings lists. Therefore, for a more stable and abridged analysis, we focused on the top 20 HEIs according to the Grand Average Rank (GAR) scores. The names and scores of the top-20 HEIs for the 9-Rankings average are presented in Fig. 2. According to the graph, Stanford University (SU) leads the top-performing HEIs with a score of 2.44. The score is the average of the nine ranking scores allocated to the HEI by the three reputed systems in the stipulated three years. Whereas the Duke University (DkU) with a 22.67 GAR score, is in the twentieth position. The mathematical model for the selection of the top hundred HEIs is given in Eq. 1.

$$\begin{aligned} &Top100HEIS = \bigcup(\forall RankEnt \cap_{Y=2017}^{2019} Top100HEIs_{R}ankEnt) \\ &\text{where} \quad RankEnt\epsilon\{ARWU, THE, andQs\} \end{aligned} \tag{1}$$

On the other side, we extracted profiles of the influential research scholars belonging to the four scientific disciplines from the GSC repository. As the GSC maintains profiles of thousands of researchers in each of the disciplines, so we limited our data retrieval to two hundred researchers for each of the disciplines. For this, we applied a threshold of hundred publications (PC = > 100) in the selection of a research scholar. The mathematical equations for the four sets { Res_{cs}, Res_{hs}, Res_{ems}, Res_{eng}} containing the researcher's profiles are given below:

$$R_{\text{cs}} = \{R_{\text{cs}} | R_{\text{cs}}(\text{PC}) \geq 100\} \tag{2}$$

$$R_{\text{hs}} = \{R_{\text{hs}} | R_{\text{hs}}(\text{PC}) \geq 100\} \tag{3}$$

$$R_{\text{ems}} = \{R_{\text{ems}} | R_{\text{ems}}(\text{PC}) \geq 100\} \tag{4}$$

$$R_{\text{eng}} = \{R_{\text{eng}} | R_{\text{eng}}(\text{PC}) \geq 100\} \tag{5}$$

$$R = R_{\text{cs}} \cup R_{\text{hs}} \cup R_{\text{ems}} \cup R_{\text{eng}} \tag{6}$$

Overall, we retrieved profiles of eight hundred researchers by taking unions of the four sets, as expressed in Eq. 6. GSC maintains various valuable statistics about the researchers and millions of their research publications. In our research analysis, we extracted various statistics including the Research Productivity (or Publication Count—PC), Research Impact (or Citation Count—CC), H-Index (HI), and Author's Research Contribution (or Author Per Paper—PP). We also retrieved the title, publication venue, and co-author(s) of the research publications published by the selected researchers. A brief description of these attributes is summarized in Table 3.

After collections of the profiles, we used the dataset to extract the four parameters—PC, CC, HI, and APP as the performance indicators of a research scholar. The selection of these parameters is based on two facts. First, in expert's opinions, these parameters are treated as the Key Performance Indicators (KPIs) in the related literature (Giovanni & Ciriaco, 2015; Sheeja et al., 2018). Second, most of the bibliographic data repositories essentially record them. An overview of the meta-data of GSC, DBLP, Microsoft Academics, and the Arnet-Miner data repositories, confirms the fact (Huang et al., 2020). The metrics themselves are also self-explanatory (S. N.K. et al., 2018). The *Publication Count* (PC) is the count of the number of publications by a researcher, it is a plain representation of the *Research Productivity* of a researcher. The *Citation Count* (CC) is a well-established metric in the academic world (Nicolas et al., 2019; Xuanyu et al., 2016). It is the count of the citations of a researchers' publications. It represents the impact or influence of a research article after its publication. The *Hirsh-Index* (HI) invented by Jorge Hirsch in 2005, is also a well-known measure to gauge both, the quantity or quality of work of a research scholar (Hirsch, 2005). Although the *Hirsh-Index* faces many critics, nevertheless, the wisdom behind the measure is very foundational; therefore, the metric in original, or its variants has been surviving since last fifteen years, in the bibliographic data repositories (Ferrara & Romero, 2013; Lutz & Daniel-Dieter, 2007).

Finally, we discuss the indicator *Author Per Paper* (APP) score, that is employed in the research analysis. The APP score is an effort to gauge the research contribution of a research scholar, especially in joint research endeavors. This measure also has become significant for the research journals and organizers of the research conferences. Usually, they inquire about the contributions of the authors, in the case of a joint research publication. Assessing the contribution of an author in a joint research effort is a complicated task. Author ranking experts have proposed various measures in the literature (Amjad et al., 2017; Daud et al., 2010; Jun et al., 2018a). Nevertheless, since our research work is focused on mapping affiliations of the influential researchers with the top-performing HEIs; therefore, we adopted the simple approach of assigning fifty percent weight to the first author and dividing the remaining fifty percent, equally among the co-authors (Dewi Ahmad & Abu Bakar, 2018; Zhang et al., 2019). The exceptional conditions like the mega-authorship or when authors are sorted alphabetically; will need exceptional weighting criteria.

After the selection of the indicators for the researcher's ranking methodology, we faced the issue of assigning weights to the parameters. The parameters' weights represent their importance in the performance evaluation. As, there is no standard weighting criterion for the academic ranking indicators, therefore they are among the debatable issues of academic rankings. We applied the well-known InfoGain (IG) (Sheeja et al., 2018; Wu et al., 2008) algorithm on the target dataset, to determines the weights of the indicators. InfoGain algorithm requires a test dataset with a class label; In our case, class (C) is the status (Influential or non-influential) of a researcher's profile. We employed the GSC default researcher's ranking criterion—the *Citation Count* (CC) to obtain the test dataset with the class label. Equation 7 represents the process of applying InfoGain to the target dataset (*R)* along with the class label.

$$\text{InfoGain}(R, C) \rightarrow w_1\text{HI}, w_2\text{CC}, w_3\text{PC}, w_4\text{APP} \tag{7}$$

The InfoGain algorithm, produced $w_1 = 36.82$, $w_2 = 30.76$, $w_3 = 25.82$, and $w_4 = 6.60$ as the weights for the indicators. GSC orders the researchers' profiles according to their *Research Impact* or *Citation Count* (CC). Nevertheless, agreeing with the fact that alone the quantity does not ensure the quality, we adopted a more comprehensive methodology to rank the research scholars. Equation 8 represents the researcher's ranking methodology with the weighted parameters.

$$ResRank(ResProf) = 36.82HI + 30.76CC + 25.82PC + 6.60APP \tag{8}$$

Where the *ResRank (ResProf)* is the function to rank a researcher's profile (*ResProf*) as influential or non-influential. As per the equation, the value of *ResRank* is the simple aggregate of the scores of the weighted indicators. The score was calculated for each researcher according to Eq. 8. Usually, the values of the selected parameters in the methodology vary significantly among the different scientific disciplines. So, the inter-discipline comparison of the researcher's profiles does not produce justifiable ranking results. Therefore, we applied the ranking methodology to the profiles within a discipline. These intra-discipline rankings can be computed using the following set of equations. Where $InfRes_{cs}$ represents the ordered list of researchers belonging to the Computer Science discipline. Similarly, the lists $InfRes_{ems}$, $InfRes_{hs}$ and, $InfRes_{eng}$ represent the Economics & Management Sciences, Health Sciences, and Engineering and Technology, respectively.

$$InfRes_{cs} \rightarrow ResRank(R_{cs}) \tag{9}$$

$$InfRes_{ems} \rightarrow ResRank(R_{ems}) \tag{10}$$

$$InfRes_{hs} \rightarrow ResRank(R_{hs}) \tag{11}$$

$$InfRes_{eng} \rightarrow ResRank(R_{eng}) \tag{12}$$

Overall, the dataset having ranked researcher's profiles of all the disciplines was developed according to Eq. 13.

$$\mathrm{InfRes} = \mathrm{InfRes}_{\mathrm{cs}} \cup \mathrm{InfRes}_{\mathrm{hs}} \cup \mathrm{InfRes}_{\mathrm{ems}} \cup \mathrm{InfRes}_{\mathrm{eng}} \tag{13}$$

The dataset produced, is the subject-specific ranking of the research scholars belonging to the four disciplines. Earlier, we also have developed the list of the HEIs that consistently retained their positions in the top-100 HEIs with outstanding performance, in the stipulated three years. Now, we have the required statistics for the fine-grained analysis of the academic rankings for *Research Productivity.* Initially, the analysis was done on the institution level, by mapping the affiliation of all the researchers in the dataset. For this, we employed the mapping function presented in Eq. 14.

$$UniRank(InfRes, TopHEIs) \tag{14}$$

The *UniRank* function takes the two datasets—*Influential Researchers* and *TopHEIs* and ranks the top-performing HEIs for the count of the influential researchers. The ranking requires mapping of the researcher-affiliation with the top-performing HEIs. In the case of multiple affiliations of an author, the first affiliation is considered, and others are ignored. The mapping highlights the HEIs having the highest number of influential researchers; alternatively, we can describe it as the HEIs ranking for the *Research Productivity* irrespective of the research disciplines. Nevertheless, we drilled down the ranking process and found the subject-specific ranking of HEIs w.r.t *Research Productivity* using the following set of equations.

Table 4 Top 20 HEIs w.r.t. the Influential Research Faculty (overall) as in December 2019

Rank	Higher Education Institution (HEI)	Acronym	Faculty Count	UniRank
1	Stanford University	SU	43	2
2	Harvard University	HU	73	1
3	Massachusetts Institute of Technology	MIT	13	13
4	University of Cambridge	UoC	18	8
5	University of Oxford	UoX	12	14
6	California Institute of Technology	CIT	2	60
7	University of Chicago	UCh	10	15
8	Princeton University	PrU	7	22
9	University of Pennsylvania	UoP	8	20
10	Swiss Federal Institute of Technology (ETH) Zurich	ETH	0	n.a
11	Yale University	YU	31	3
12	University College London	UCL	5	26
13	Columbia University	ClU	28	4
14	Imperial College of Science Technology and Medicine London	ICL	22	6
15	Cornell University	CrU	7	22
16	University of California at Berkeley	UCB	15	10
17	Johns Hopkins University	JHU	4	32
18	The University of California at Los Angeles	UCLA	4	32
19	University of Michigan	UMi	20	7
20	Duke University	DkU	4	32

$$UniRank_{\mathrm{cs}}(InfRes_{\mathrm{cs}}, TopHEIs) \tag{15}$$

$$UniRank_{\mathrm{hs}}(InfRes_{\mathrm{hs}}, TopHEIs) \tag{16}$$

$$UniRank_{\mathrm{ems}}(InfRes_{\mathrm{ems}}, TopHEIs) \tag{17}$$

$$UniRank_{\mathrm{eng}}(InfRes_{\mathrm{eng}}, TopHEIs) \tag{18}$$

We proceeded with the fine-grained analysis on the same pattern and developed the HEIs rankings for the count of their *Research Publications* (PC) and *Research Impact* or *Citation Count* (CC). The results are presented in section "Results and discussion".

Experiments

We acquired the required data from the data sources introduced in section "Data sources", using multiple techniques. To obtain the list of top-100 HEIs with consistently outstanding performance in the stipulated three years, we retrieved the HEI's rankings from official web portals of ARWU, THE, and QS academic ranking systems. The profiles of influential research scholars, having a high Citation Count (CC) or research impact, were retrieved from the GSC data repository. To obtain the necessary data about the HEIs and countries,

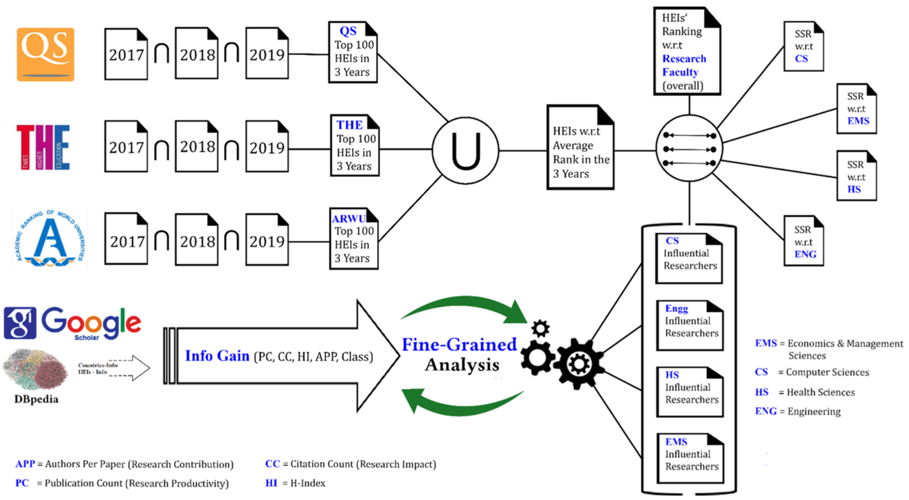

Fig. 3 Workflow of the fine-grained analysis

we employed the DBpedia data repository. The data in the DBpedia was processed through its SPARQL Endpoint[12] using the SPARQL queries (Holst & Höfig, 2013). DBpedia provided the data about the HEIs (i.e., their formal name, location, country, foundation year, URL, etc.) and countries (i.e., their region, HDI index, total population etc.). Since the DBpedia dataset is not curated for the HEIs' information, therefore in some cases, we verified the data from websites of the HEIs. Finally, we mapped the researchers' affiliation with the top-performing HEIs to obtain the fine-grained ranking of HEIs w.r.t. the research productivity. The list of the top twenty HEIs w.r.t. their research Faculty (overall) is presented in Table 4. For a quick understanding, please consult the process diagram presented in Fig. 3. The diagram depicts various steps of data retrieval and analysis at the abstract level.

The fine-grained analysis of the data would help in understanding the causes of the different ranking scores of HEIs, as is suggested in Jo (2019); Meymandpour & Davis, 2016). The holistic ranking scores published by various ranking systems face the criticism termed as the "*Misinterpretation of the academic performance*" because of such discrepancies. We can minimize such discrepancies by the subject-specific rankings or even better by ranking HEIs w.r.t. various ranking dimensions such as the *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.*

Data pre-processing

After the data acquisition, we selected the top-100 HEIs with outstanding performance according to the GAR score, in the stipulated three years (2017, 2018, and 2019). The selection method is modeled in Eq. 1. On the other hand, we selected profiles of the influential research scholars having considerable publication count ($\mathrm{PC} \geq 100$), belonging to the four disciplines. The selection criteria are modeled in the set of equations labeled as

[12] https://dbpedia.org/sparql.

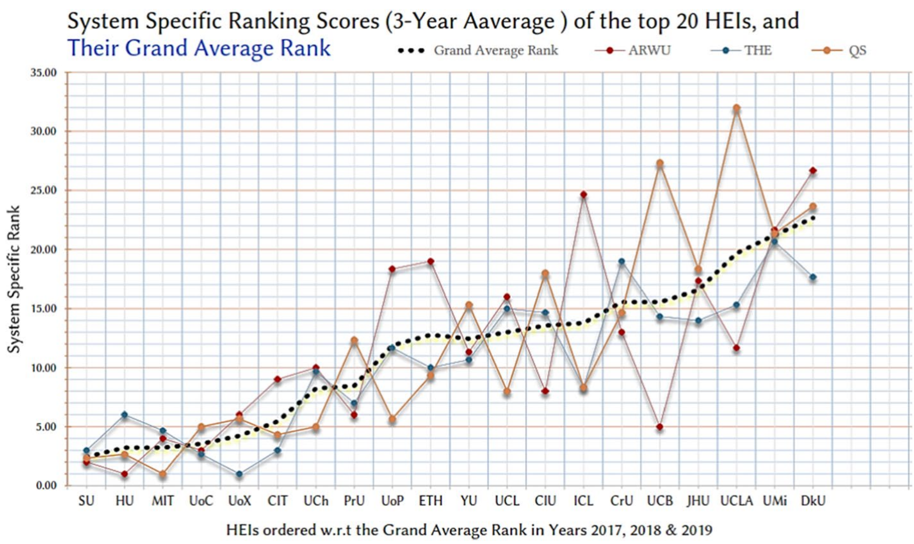

Fig. 4 Top 20 HEIs w.r.t. the 3-year Average Rankings by ARWU, THE, QS, and GAR score

Eqs. 2 to 5. Then, we accumulated the profiles by the union operation presented in Eq. 6. We applied the researcher's ranking methodology on the profiles for evaluation of their influential status. The mathematical expression of the methodology is presented in Eq. 7.

Results and discussion

The Internet offers various data repositories containing data related to academics, and such data sources are continually increasing. Nevertheless, there is no purposely built benchmark data source for the academic rankings. We employed the DBpedia and GSC repositories because of their high relevance with the intended analysis. In the future, the fine-grained analysis strategy would employ more relevant and curated data sources.

Analysis of the deviation in the rankings

The first motive of the fine-grained strategy was the observation of the considerable differences among the ranking scores allocated by ARWU, THE, and QS, to an HEI even in a similar year. To analyze the difference, we compared the 3-year average of the ranking scores allotted to the HEIs by each of the systems with the GAR score. The resultant comparison is presented in Fig. 4. In the figure, the solid-color lines represent the 3-year averages of the ranking scores allotted to the HEIs by the three systems. Whereas, the dotted line represents the GAR score, attained by the HEIs. According to the plot, the deviation from the GAR score is minimal for only a few top-performing HEIs. Nevertheless, the deviation becomes significant for the HEIs with higher ranking scores. For example, the 3-year average score allotted to the *Imperial College London* (ICL) by QS and THE is 8, while the same institution is allocated a 3-year average score equals to 25 by the ARWU. ICL being among the top ten world-class institutions would be reluctant in accepting such a large difference.

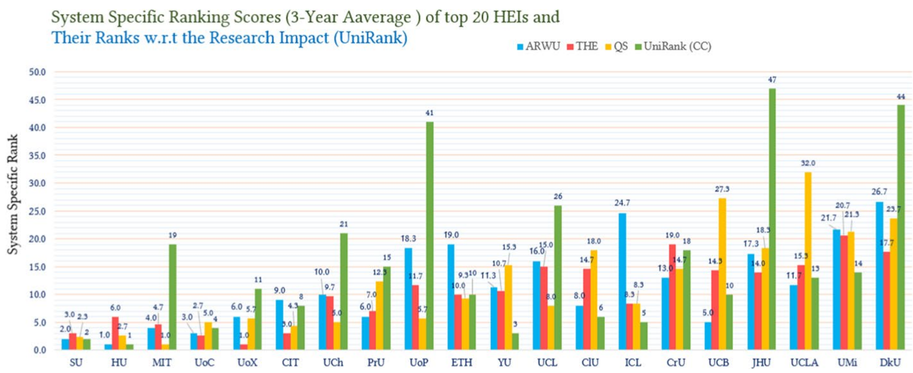

Fig. 5 Top 20 HEIs w.r.t. the GAR score in years 2017–2019, and Research Impact

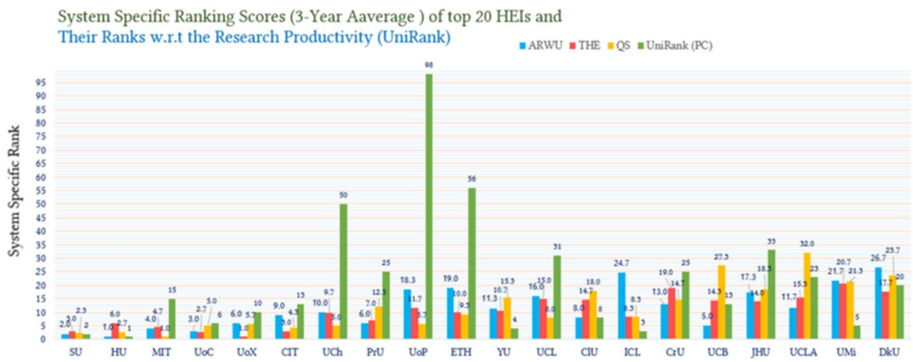

Fig. 6 Top 20 HEIs w.r.t. the GAR score in years 2017–2019, and Research Productivity

Similarly, the QS and ARWU have significant differences (22 and 19 respectively) in the 3-year average ranking scores for the two HEIs—the *University of California at Berkley* (UCB) and the *University of California at Los Angles* (UCLA).

Analysis of the research impact

Gauging the research impact of the research faculty affiliated with an HEI is also among the common objectives of the ranking methodologies of the reputed academic ranking systems. Commonly, the metrics like Citation Count and H-Index are employed for this purpose.

Keeping in view the information available in the GSC data repository, we extracted and accumulated the Citation Count (CC) attained by the Influential Research Faculty (IRF) affiliated with the top-performing HEIs. Figure 5 presents the ranking of the HEIs w.r.t. the Research Impact or Citation Count (CC) (Fig. 6). In the figure, the green color bars represent the HEIs ranks w.r.t. the research impact—UniRank (CC), whereas the other color bars represent the 3-year average rankings by ARWS, THE, and QS.

According to the graph, the research impact is in consonant with the ranking scores of only a few (hardly fifteen) top-performing HEIs. Afterward, its value differs significantly from the 3-year average ranking scores allocated by the three systems. For example, the UniRank scores attained by Johns Hopkins University (JHU) and Duke University (DkU) are 47 and 44, respectively. Nevertheless, these HEIs achieved very different ranking scores by ARWU, THE, and QS. Such differences shake the confidence of the audience in the HEIs rankings, and they would require the fine-grained analysis of the holistic ranking scores published in the rankings lists of the reputed ranking systems.

For brevity, we presented a comparison of top-20 HEIs. In general, the difference in ranking scores becomes more significant for the HEIs having a rank greater than thirty-five, in ranking lists of any of the three ranking systems. That is why the electronic and print media advertise more condensed academic ranking lists labeled as the “Top 25”, “Top 50” or “Top 100” world-class HEIs with outstanding performance. After observing such significant differences, the end-user would naturally hesitate in believing an academic ranking system. The differences also provide a clue for criticizing a ranking system in terms of favoritism and commercialization. Such discrepancies increase the need for fine-grained analysis of the academic data. Heading towards the more transparent and fine-grained analysis is inevitable in the ensuing years.

Analysis of the research productivity

While gauging the research productivity, usually the metrics like the number of field-medals, registered patents, and publication count are employed. We also extracted and accumulated the Publication Count (PC) recorded against the influential researchers affiliated with the top-performing HEIs. This ranking is termed as the HEIs ranking w.r.t. Research Productivity (UniRank). The list of the top twenty HEIs, w.r.t. the research publications are presented in Table 5.

For a better understanding, the rankings are visualized in Fig. 8. In the figure, the green color bars represent the ranking w.r.t. the Research Productivity, whereas the other colored bars represent the 3-years average ranking scores of HEIs that were allocated by ARWU, THE, and QS.

Interestingly, as per the graph, the HEIs ranking w.r.t. the research productivity agrees with the ranking scores of top-performing HEIs. The good coincidence provides a clue of the availability of the statistics related to research publications by the influential researchers affiliated with the top-performing HEIs. The constantly flourishing and publicly available bibliographic data repositories on the Internet made the analysis possible. The same level of availability of the other metrics related to the research productivity would help in developing a more transparent, fine-grained, and publicly verifiable academic ranking.

While analyzing the research productivity of the HEIs, we also extracted and accumulated the subject or discipline-specific Publication Count (PC). This count was recorded against the Influential Research Faculty (IRF) affiliated with the top-performing HEIs in a specific discipline. The subject-specific ranking by UniRank is based on the statistics. In Fig. 7, the subject-specific rank of the top 20 HEIs, is represented by the dark blue, yellow, blue, and green bars.

We determined the ranks of HEIs in the four disciplines w.r.t. the research productivity. The lists of the top five research scholars in the four scientific disciplines are presented in the four tables—Table 6 Computer Science (CS), Error! Reference source not found. Economics and Management Sciences (EMS), Error! Reference source not found (Tables 7, 8, 9, 10). Health Sciences (HS), and Table 11 Engineering (ENG). The subject-specific rankings are collectively visualized in Fig. 7. An overview of the graph reveals the fact

Table 5 Top 5 Researchers in Computer Sciences w.r.t. the UniRank methodology (Nov 2019)

Researcher	Affiliation	H-Index	Publication Count	Citation Count	APP Score
Robert Tibshirani	Stanford University	142 (162)*	967 (1019)	285,598 (382,869)	364.93
Anil K. Jain	Michigan State University	182 (198)	995 (1017)	192,561 (225,588)	393.76
Jiawei Han	University of Illinois Urbana-Champaign	168 (183)	1446 (1564)	170,320 (157,743)	488.09
Federico Calzolari	Scuola Normale INFN CERN	120 (137)	2722 (2841)	83,485 (107,508)	2637.54
Thomas S. Huang	University of Illinois, Urbana-Champaign	143 (165)	1993 (2092)	150,248 (176,152)	718.22

*Current statistics (June 2021) are inside the brackets

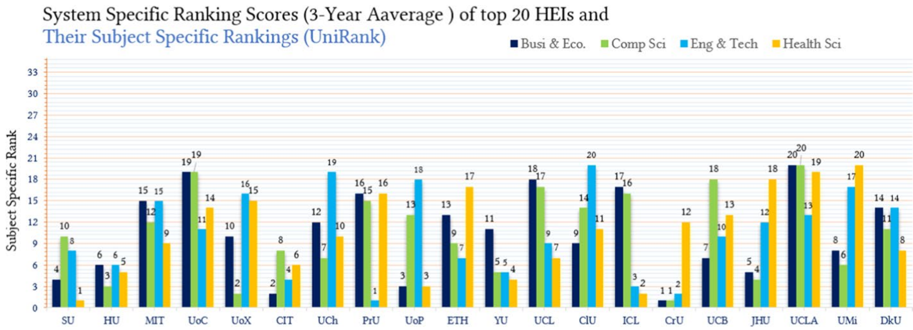

Fig. 7 Subject-Specific Ranks of top 20 HEIs, in years 2017–2019

Table 6 Top 20 HEIs w.r.t. Research Publications by their influential researchers, as in Nov 2019

Rank	Higher Education Institution (HEI)	Acronym	Total Publications	UniRank Score
1	Stanford University	SU	25,298	2
2	Harvard University	HU	53,598	1
3	Massachusetts Institute of Technology	MIT	4757	15
4	University of Cambridge	UoC	9422	6
5	University of Oxford	UoX	6351	10
6	California Institute of Technology	CIT	417	145
7	University of Chicago	UCh	2035	50
8	Princeton University	PrU	3639	25
9	University of Pennsylvania	UoP	1084	98
10	Swiss Federal Institute of Technology (ETH) Zurich	ETH	0	n.a
11	Yale University	YU	13,444	4
12	University College London	UCL	2952	31
13	Columbia University	ClU	7999	8
14	Imperial College of Science Technology and Medicine London	ICL	17,238	3
15	Cornell University	CrU	3602	26
16	University of California at Berkeley	UCB	5018	13
17	Johns Hopkins University	JHU	2946	33
18	The University of California at Los Angeles	UCLA	3763	23
19	University of Michigan	UMi	12,114	5
20	Duke University	DkU	4196	20

that even the top ten HEIs have significantly different ranking scores in the four scientific disciplines. For example, Stanford University (SU) obtained the first position according to the GAR score; Nevertheless, its subject-specific rankings are 4th, 10th, 8th, and 1st for the Business & Economics, Computer Sciences, Engineering and Technology, and Health Sciences, respectively.

According to the plot, the difference between the IRF specific rank (UniRank) and other rankings scores is nominal for only a few top-performing HEIs. Nevertheless, the difference

Table 7 Top 5 Researchers in Economics and Management w.r.t. the UniRank methodology (Nov 2019)

Researcher	Affiliation	H-Index	Publication Count	Citation Count	APP Score
Michael E. Porter	Harvard University	175 (181) *	1948 (2142)	479,415 (514,877)	1699.80
Robert Kaplan	Harvard University	94 (100)	657 (833)	165,019 (192,672)	4079.41
Lawrence Summers	Harvard University	174 (184)	1523 (1668)	142,894 (166,098)	1855.90
Kenneth J. Arrow	Stanford University	149 (162)	1857 (2037)	205,116 (237,434)	1256.44
James Heckman	University of Chicago	164 (173)	1063 (1158)	190,164 (228,053)	571.73

*Current statistics (June 2021) are inside the brackets

becomes significant for HEIs with higher ranking scores. For example, the IRF score attained by Princeton University (PrU) is 20, whereas the university is attaining the ranking scores 6, 7, and 12 by ARWU, THE, and QS, respectively. Similarly, we can observe other HEIs having significant differences including the Johns Hopkins University (JHU) and the University of California at Los Angles (UCLA). The differences become more significant for the HEIs with higher ranking scores. Usually, the holistic ranking scores published by various reputed academic ranking systems face criticism terms as "*Misinterpretation of the academic performance*". Such discrepancies can be minimized by the subject-specific rankings or even better by ranking HEIs w.r.t. various ranking dimensions such as the *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.*

Analysis of the Influential Research Faculty

In experts' opinions (Jamil, 2009; Lokman, 2020), the prime objective of an HEI is the generation of new knowledge. That is why most of the global academic ranking systems essentially gauge the Research Productivity and Research Impact of HEIs. We also analyzed these ranking dimensions. We mapped the affiliation of the influential researchers (in all the disciplines) with the top-performing HEIs according to the GAR score. A part of the mapping results is presented in Fig. 8. In the figure, the bars with green color represent the HEIs ranking w.r.t. the presence of the Influential Research Faculty (IRF) in the four scientific disciplines. Whereas the other bars represent the 3-year average of the ranking scores allotted to the HEIs, by each of the three systems—ARWU, THE, and QS.

Considering the context in the Rankings

One of our data sources was the DBpedia data repository. We extracted various meaningful information about the countries and HEIs including the base country of an HEI, its exact location, city, year of establishment, etc. Availability of the year of establishment of HEIs helped us in assessing their academic age. A part of the result is presented in Fig. 9. An overview of the graph reveals the fact that all the top-20 HEIs with outstanding performance have academic more than a hundred years. It means achieving outstanding status at the world level, requires decades of consistent effort.

Table 8 Top 20 HEIs with their rank in the different disciplines w.r.t. the Influential Research Faculty

Rank	Higher Education Institution	Acronym	HS	CS	E&T	BE	GAR Score
1	Stanford University	SU	38 (1)	10 (1)	8 (3)	12 (6)	2.44
2	Harvard University	HU	1 (18)	0 (n.a)	2 (21)	32 (1)	3.22
3	Massachusetts Institute of Technology	MIT	14 (4)	0 (n.a)	2 (21)	9 (10)	3.22
4	University of Cambridge	UoC	1 (18)	0 (n.a)	0 (n.a)	4 (18)	3.56
5	University of Oxford	UoX	0 (n.a)	1 (26)	4 (7)	5 (17)	4.22
6	California Institute of Technology	CIT	0 (n.a)	1 (26)	1 (39)	0 (n.a)	5.44
7	University of Chicago	UCh	0 (n.a)	0 (n.a)	0 (n.a)	10 (8)	8.22
8	Princeton University	PrU	1 (18)	1 (26)	6 (5)	0 (n.a)	8.44
9	University of Pennsylvania	UoP	0 (n.a)	0 (n.a)	0 (n.a)	7 (14)	11.89
10	Swiss Federal Institute of Technology (ETH) Zurich	ETH	0 (n.a)	0 (n.a)	0 (n.a)	0 (n.a)	0 (n.a)
11	Yale University	YU	1 (18)	0 (n.a)	1 (39)	13 (4)	12.44
12	University College London	UCL	2 (8)	2 (13)	2 (21)	0 (n.a)	13.00
13	Columbia University	ClU	19 (2)	4 (3)	7 (4)	15 (2)	13.56
14	Imperial College of Science Technology and Medicine London	ICL	0 (n.a)	1 (26)	2 (21)	0 (n.a)	13.78
15	Cornell University	CrU	0 (n.a)	3 (4)	4 (7)	0 (n.a)	15.56
16	University of California at Berkeley	UCB	1 (18)	3 (4)	0 (n.a)	12 (6)	15.56
17	Johns Hopkins University	JHU	2 (8)	1 (26)	2 (21)	0 (n.a)	16.56
18	The University of California at Los Angeles	UCLA	2 (8)	1 (26)	1 (39)	0 (n.a)	19.67
19	University of Michigan	UMi	1 (18)	4 (3)	13 (2)	0 (n.a)	21.22
20	Duke University	DkU	38 (1)	1 (26)	1 (39)	0 (n.a)	22.67

Table 9 Top 5 Researchers in Health Sciences w.r.t. the UniRank methodology (Nov 2021)

Researcher	Affiliation	H-Index	Publication Count	Citation Count	APP Score
Frank B. Hu	Harvard University	234 (278)*	2001 (2363)	226,525 (360,114)	475.84
Dr. Joann E. Manson	Harvard University	266 (298)	1500 (2822)	273,626 (369,557)	346.04
Daniel Levy	National Heart Lung and Blood Institute	207 (245)	1976 (2101)	213,572 (319,635)	535.98
Guido Kroemer	Université de Paris Hôpital Européen	227 (247)	1597 (1885)	225,593 (282,207)	425.75
Harlan Krumhold	Yale University	191 (214)	2001 (2417)	177,379 (252,018)	511.20

*Current statistics (June 2021) are inside the brackets

Table 10 Top 20 HEIs w.r.t. the Research Impact (Citation Count) by their influential researchers, as in Nov 2019

Rank	Higher Education Institution (HEI)	Acronym	Citation Count	UniRank
1	Stanford University	SU	3,282,264	2
2	Harvard University	HU	7,740,062	1
3	Massachusetts Institute of Technology	MIT	327,980	21
4	University of Cambridge	UoC	1,190,102	6
5	University of Oxford	UoX	579,673	11
6	California Institute of Technology	CIT	16,602	157
7	University of Chicago	UCh	285,775	27
8	Princeton University	PrU	373,859	19
9	University of Pennsylvania	UoP	161,276	50
10	Swiss Federal Institute of Technology (ETH) Zurich	ETH	0	n.a
11	Yale University	YU	1,529,522	3
12	University College London	UCL	294,022	26
13	Columbia University	ClU	1,209,813	5
14	Imperial College of Science Technology and Medicine London	ICL	1,024,013	7
15	Cornell University	CrU	310,595	23
16	University of California at Berkeley	UCB	788,457	10
17	Johns Hopkins University	JHU	231,683	35
18	The University of California at Los Angeles	UCLA	511,565	13
19	University of Michigan	UMi	416,320	16
20	Duke University	DkU	179,746	44

The data about the geographical regions and countries of HEIs can be extracted from the DBpedia repository. It also provides information about the Human Development Index (HDI) of a country. During the data analysis, we explored the relationship between HDI of a country and the count of its top-performing HEIs, according to the GAR score. The results of the analysis are presented in Fig. 10. Interestingly, according to the trend, one can argue that the outstanding human prosperity of a country does not ensure its outstanding share in scientific research and innovation. Nevertheless, the shares of the United States of America and the United Kingdom are advocating the need for human prosperity for better performance.

We also explored the relationship between the Gross Domestic Product (GDP) of a country and the number of top-performing HEIs (according to the GAR score) owned by it. Figure 11 presents a part of the result of the analysis. Here, one can argue that the high GDP of a country indicates its vibrant economy, which nurtures research and innovation.

Such countries usually have heavy industries that require a highly qualified workforce through outstanding HEIs. The United States of America, the United Kingdom, China, Japan, and Germany are dominant in sharing the HEIs having the top-performance. Moreover, the HEIs also enjoy conducive settings in terms of their base cities, and countries having stable political and economic conditions.

Similarly, using the data about the base countries, we also identified the countries having a considerable number of HEIs in their different cities. Quite understandably, the top ten countries having HEIs with outstanding performance, are the most stable and

Table 11 Top 5 Researchers in Engineering and Technology w.r.t. the UniRank methodology (Nov 2019)

Researcher	Affiliation	H-Index	Publication Count	Citation Count	APP Score
Nicholas A. Peppas	UT Texas Austin	167 (187)*	2001 (2177)	130,168 (159,570)	1040.41
Yang	University of California Los Angeles	149 (161)	3000 (3000 +)	101,873(122,185)	816.65
Sercan Sen	Hacettepe University	164 (187)	2101 (2785)	125,329 (159,671)	572.37
Lev Dudko	Lomonosov Moscow State University	164 (186)	1571 (1963)	144,708 (188,347)	399.32
Rob Knight	University of California San Diego	166 (199)	675 (890)	167,776 (257,855)	146.82

* Current statistics (June 2021) are inside the brackets

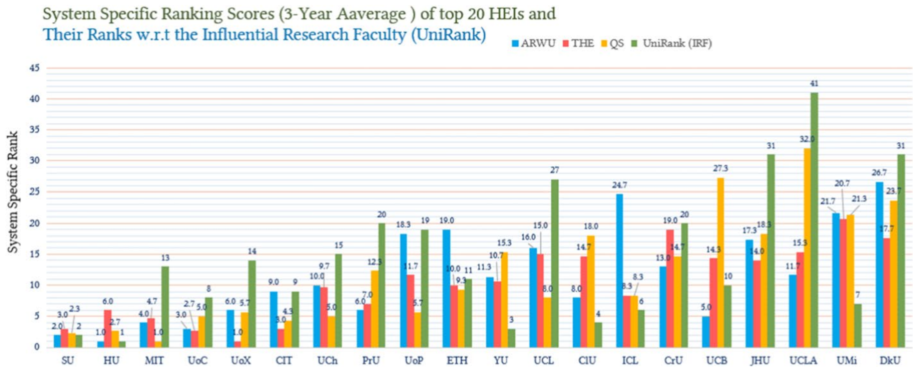

Fig. 8 Top 20 HEIs w.r.t. the GAR score in years 2017–19, and Influential Research Faculty (IRF)

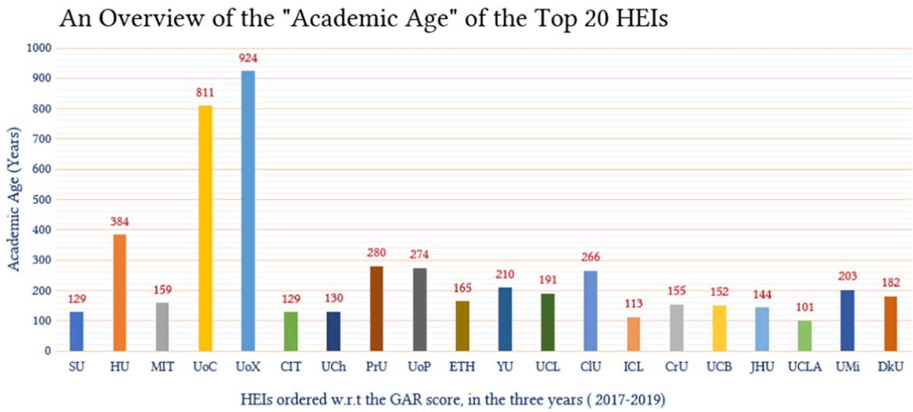

Fig. 9 An overview of the "Academic Age" of top-performing HEIs

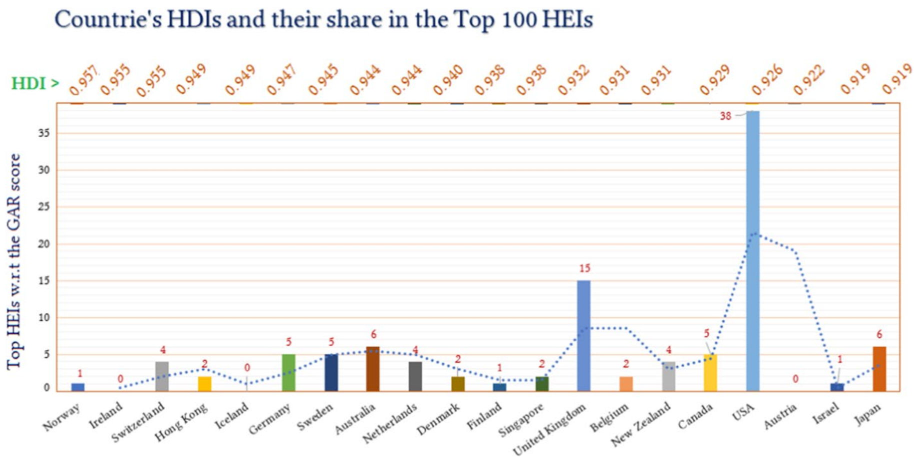

Fig. 10 Top 20 Countries w.r.t. the HDI (UNDP, 2019) vs Top HEIs w.r.t. the GAR Score

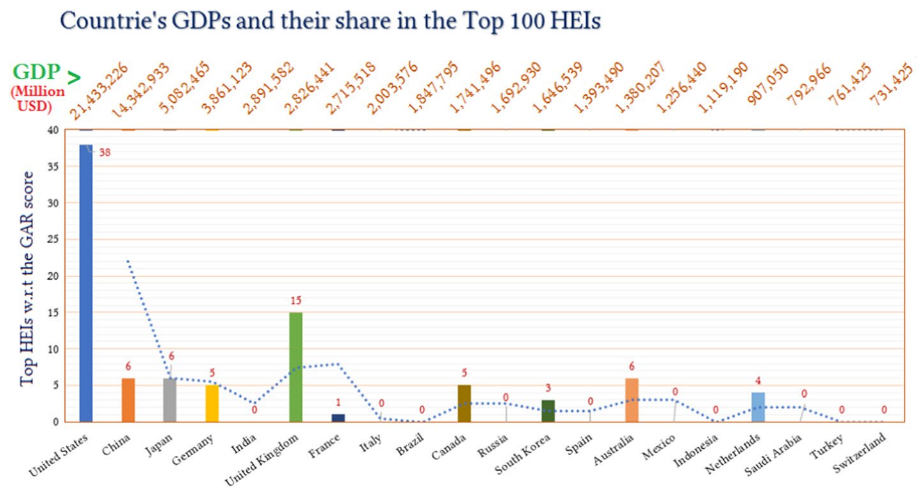

Fig. 11 Top 20 Countries w.r.t. the GDP Per Capita (UN, in 2019) vs Top HEIs w.r.t. the GAR Score

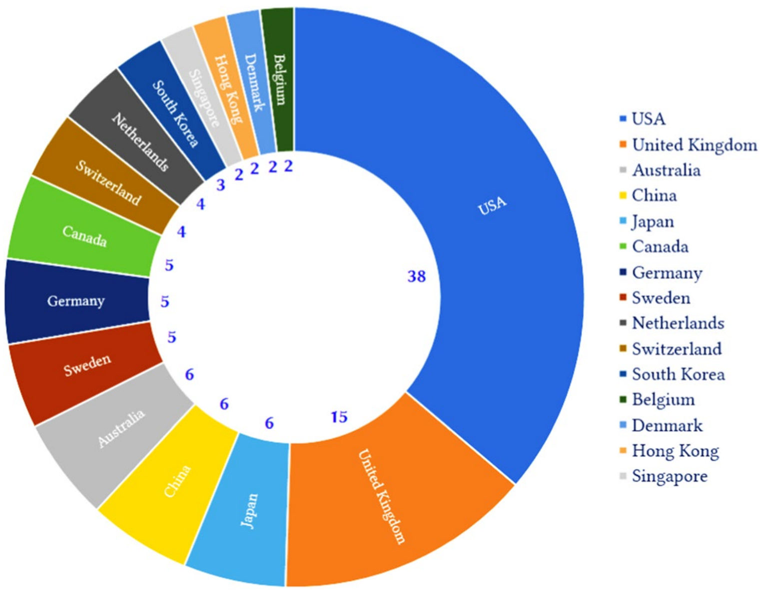

Fig. 12 Top 15 Countries with their share in the Top Perfuming HEIs w.r.t. the GAR Score

prosperous courtiers. Figure 12 represents a part of the analysis. Governments of the developed countries are cognizant of the role of HEIs in the advancement of their socio-economic development.

Thus, they nurture research, innovation, and commercialization through the institutions. Some academic experts also criticize the ranking process of "Global HEIs" for not caring about these subtle differences while allocating ranking scores to the HEIs. In (Qureshi et al., 2021), we discussed the argument that comparisons of HEIs having very different sizes, budgets, and locations, etc. also raise the question of the validity of the ranking process.

It is worth mentioning that the leading academic ranking systems are heading towards the more fine-grained analysis in terms of their academic age and geographic location. The ranking lists having titles like "Top 50 Under 50" and "Top 25 HEIs in Asia" are the outcome of such fine-grained analysis.

Conclusions and future work

This research effort was aimed at enhancing the creditability of the academic ranking process by fine-grained analysis of the academic ranking data. Using a newly proposed researcher's ranking methodology, we mapped the affiliation of the influential researchers with the top-performing HEIs, to explore their research productivity and research impact. We also demonstrated the effectiveness of the proposed data analysis strategy through the publicly verifiable data repositories—the DBpedia and Google Scholar Citations. Here are the conclusions and future implications of the analysis.

Conclusions

HEIs' ranking systems significantly have improved their quality during the past fifteen years. Currently, they are more informative and user-friendly than they were a decade ago. Nevertheless, there are considerable differences among the ranking lists published by the reputed academic ranking systems. The fine-grained analysis w.r.t. the various ranking dimensions (such as the *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.*) would reveal the factors, causing significant differences among the rankings scores of HEIs.

The holistic ranking score, even in the subject-specific rankings, hides many valuable aspects of a research institution. Whereas the fine-grained analysis makes the rankings score more representative, and the ranking process less controversial. At the sub-discipline level, various HEIs perform remarkably well, although they are not visible on the top, in the lists of reputed academic ranking systems.

The fine-grained analysis also suggests considering the context of HEIs (i.e., financial condition of the base county, Academic Age, Language, and Size) while comparing the HEIs, even in the subject-specific rankings. Ignoring the context would lead to unfair comparisons.

Future work or implications

In the future, we shall focus on improving the accuracy of the fine-grained analysis, by employing more relevant public data sources including the bibliographic data maintained by Association for Computing Machinery (ACM) and Scopus, Microsoft Academics, and LinkedIn repositories, etc. Such data sources would reveal the new dimensions of HEIs' academic performance. As per the pattern of this research work, the rankings of other scientific disciplines would be analyzed w.r.t. the various ranking dimensions. The reputed international rankings would be the first to expand the limited picture of global higher education.

The quantitative academic ranking indicators employed in the analysis were extracted from openly accessible data repositories. These "quantitative" and publicly verifiable data sources make the ranking results reproducible and increase the confidence of the academic stakeholders. The diversity in the underlying data helps in a comprehensive understanding of academic performance.

The fine-grained analysis of the data related to global HEIs also demands a purposely built, publicly accessible electronic data source having quantitative ranking indicators. We can produce more transparent, reliable, and less controversial academic rankings by employing the aspired data source. Such a global data repository would help in better academic planning, management, monitoring, and analysis.

References

Federkeil (2011). Rankings and quality assurance in higher education. *Higher Education in Europe,* 33(2), 219–231,. http://dx.doi.org/https://doi.org/10.1080/03797720802254023.

Qureshi, M. S., Daud, A., Hayat, M. K., & Afzal, M. T. (2021). OpenRank—A novel approach to rank universities using objective and publicly verifiable data sources. *Library Hi-Tech,* 39(1).

Lokman, I. M. (2020). Highly prestigious international academic awards and their impact on university rankings. *Quantitative Science Studies, 1*(2), 824–848.

Abramo, G., Andrea, D. C., & Soldatenkova, A. (2016). The ratio of top scientists to the academic staff as an indicator of the competitive strength of universities. *JournalofInformetrics, 10*, 596–605.

Fredrik, N. P., & Gunnar, S. (2016). How can differences in international university rankings be explained? *Scientometrics, 109*(3), 2263–2278.

Lin, C.-S., Huang, M.-H., & Chen, D.-Z. (2013). The Influences of Counting Methods on University Rankings Based on Paper Count and Citation Count. *Journal of Informetrics, 7*, 611–621. https://doi.org/10.1016/j.joi.2013.03.007

Jo, A. O. (2019). The "Dark Side" of Academics? Emerging Issues in the Gaming and Manipulation of Metrics in Higher Education. *The Review of Higher Education, 42*(3), 859–877.

Henk, F., & Moed (2016). A critical comparative analysis of five world university rankings, *Scientometrics,* 110(2), 967–990.

Friso, S., Cameron, N., Chun-Kai, H., & Paul, G. (2020). A Longitudinal analysis of university rankings. *Quantitative Science Studies, 1*(3), 1109–1135.

Aguillo, I. F., & Bar-Ilan, J. (2010). Comparing University Rankings. *Journal of Scientometrics, 85*, 243–256.

Garcia, R., & Herrera, F. (2013). An Insight into the Importance of National University Rankings in an International Context. In *ISSI Conference*, Vienna.

Huang, C.-K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies, 1*(2), 445–478.

Ortega, J. L. (2014). Influence of co-authorship networks in the research impact Ego network analyses from Microsoft Academic Search. *Journal of Informetrics, 8*, 728–737.

Yang, B., Ronald, R., Shuiqing, H., & Sulan, Y. (2015). Do first rate scientists work at first rate organizations? *Malaysian Journal of Library & Information Science, 20*(1), 47–60.

Giovanni, A., & Ciriaco, A. D. (2015). Ranking research institutions by the number of highly-cited articles per scientist. *Journal of Informetrics, 9*, 915–923.

Mingers, J., O'Hanley, J. R., & Okunola, M. (2017). Using Google Scholar institutional level data to evaluate the quality of university research. *Scientometrics, 113*, 1627–1643.

Jabnoun, N. (2015). The influence of wealth, transparency, and democracy on the number of top ranked universities. *Quality Assurance in Education, 23*(2), 108–122.

Sheeja, N. K., Mathew, S., & Cherukodan, S. (2018). Impact of scholarly output on university ranking,". *Global Knowledge, Memory and Communication, 67*(3), 154–165.

Jun, Z., Bo, X., Jiaying, L., Amr, T., & Al, Z. (2018). PePSI: Personalized Prediction of Scholars' Impact in Heterogeneous Temporal Academic Networks, *IEEE ACCESS*.

Koen, F., Gaston, J. H., & Jarno, H. (2017). What drives university research performance? An analysis using CWTS Leiden Ranking data. *Journal of Informetrics,* pp. 859–872.

Uslu, B. (2020). A path for ranking success: What does the expanded indicator-set of international university rankings suggest? *Higher Education, 80*, 949–972.
Maxime, M., & Alex, P. J. (2018). Engineering the global university rankings: Gold standards, Limitations and Implications. *IEEE Acess,* vol. 6
Boulton, G. (2011). University Rankings: Diversity, Excellence and the European Initiative. *Science Direct Procedia Social and Behavioral Sciences, 13*, 74–82.
Luis, O. (2014). Influence of co-authorship networks in the research impact: Ego network analyses from Microsoft Academic Search. *Journal of Informetrics, 8*(3), 728–737.
Heath, B. C., & Lee, T. B. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22.
Meymandpour, R., & Davis, J. G. (2016). A semantic similarity measure for linked data: An information content-based approach. *Knowledge-Based Systems*.
Holst, T., & Höfig, E. (2013). Investigating the relevance of linked open data sets with SPARQL queries. In *37th Annual Computer Software and Applications Conference Workshops*, Berlin Heidelberg, https://doi.org/10.1109/COMPSACW.2013.31
Huang, C.-K., Cameron, N., Chloe, B.-K., Richard, H., Lucy, M., Katie, W., & Alkim, O. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies, 1*(2), 445–478.
Nicolas, R.-G., Daniel, T.-S., Enrique, H.-V., & Domingo, D. (2019). Mining university rankings: Publication output and citation impact as their basis. *Research Evaluation, 28*(3), 232–240.
Xuanyu, C., Yan, C., & R. L. K.J. (2016) A data analytic approach to quantifying scientific impact. Journal of Informetrics, 10(2): 471–484
Hirsch, J. E. (2005). An index to quantify an individual's scientific research
Ferrara, E., & Romero, A. (2013). Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index. *Journal of the American Society for Information Science and Technology, 64*(11), 2332–2339.
Lutz, B., & Daniel-Dieter, H. (2007). What do we know about the h-index? *Journal of the American Society for Information Science and Technology, 59*(9), 1381–1385.
Daud, A., Li, J., Zhou, L., & Muhammad, F. (2010). Temporal expert finding through generalized time topic modeling. *Knowledge-Based Systems, 65*(9), 70–81.
Amjad, T., Ding, Y., Xu, J., Zhang, C., Tang, J., Song, M., & Daud, A. (2017). Standing on the shoulders of giants. *Journal of Informetrics, 11*, 307–323.
Zhang, F., Bai, X., & Lee, I. (2019). Author impact: evaluations, predictions, and challenges. *IEEE Access,* pp. 125–138.
Dewi Ahmad, W., & Abu Bakar, A. (2018). Classification models for higher learning scholarship award decisions. *Asia-Pacific Journal of Information Technology and Multimedia,* 7(2), pp. 131–145.
Wu, X., Kumar, V., & Quinlan, J. R. (2008). Top 10 algorithms in data mining. *Knowledge Informaton System, 14*, 1–37.
Jamil, S. (2009). The challenge of establishing world-class universities., WashingtonDC: TheWorldBank. ,ISBN978–0–8213–7865–6.

View publication stats