Linked Open Data Based Ranking of the Global Higher Education Institutions

Ph.D. Thesis

By

Muhammad Sajid Qureshi

99-FBAS/PHDCS/F13

Supervisor

Dr. Ali Daud

Co-Supervisor

Dr. Muhammad Tanvir Afzal

Department of Computer Science & Software Engineering

Faculty of Basic and Applied Sciences

International Islamic University, Islamabad.

2022

Data Based

Information retrieval - computer pro

software engineering

A dissertation submitted to the Department of Computer Science & Software Engineering, International Islamic University, Islamabad as a partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** in Computer Science.

INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD
FACULTY OF COMPUTING
DEPARTMENT OF COMPUTER SCIENCE & SOFTWARE ENGINEERING

Date: 23-11-2022

Final Approval

It is certified that we have read this thesis, entitled "Linked Open Data Based Ranking of the Global Higher Education Institutions" submitted by **Mr. Muhammad Sajid Qureshi Registration No. 99-FBAS/PhDCS/F13**. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for the award of the degree of PhD. in Computer Science

Committee

External Examiner:
Prof. Dr. Arif ur Rehman,
Professor,
Bahria University,
Islamabad.

External Examiner:
Dr. HIkmat ullah Khan,
Assoicate Professor.
COMSATS Wah Cant.

Internal Examiner:
Dr. Syed Muhammad Saqlain.
Assistant professor
Department of Computer Science & Software Engineering
FoC, IIUI

Supervisor:
Dr. Anwar Ghani,
Lecturer on behalf of supervisor
Department of Computer Science & Software Engineering
FBAS, IIUI

Co-Supervisor:
Dr. Muhammad Tanvir Afzal.
Professor
NAMAL Institute, Mianwali

Acknowledgments

I simply bow my head before Allah Subhanahu-wa-Tahala, for giving me faith in my abilities and enabling me to accomplish this research work. He showered upon me, His unremitting blessings throughout my life.

The unparalleled self-sacrifices of my parents always inspired me to put my best efforts into every challenge of my life. Their support and prayers had been a great asset in my career. My elder brother remained a source of motivation, especially in the beginning and closing phases of the degree. I am also thankful to my elder sister, for her cooperative role in the home, during my studies. Acknowledgment is also due to my dear wife – Samina, for her interest, encouragement, and support in the research work. She came into my life during the last years of my Ph.D. studies; Nevertheless, her sincere efforts substantially catalyzed my work progress.

The humble accomplishment of this thesis could not be successful without the contribution of many individuals, to whom I must express my appreciation and gratitude.

Dr. Ali Daud – my research supervisor, especially guided and encouraged me at the beginning of the research work. Later, although he moved to Saudi Arabia, nevertheless, his prompt responses were always there for my queries. I am thankful to him for his contribution to the research activity. Dr. Muhammad Tanvir Afzal – the co-supervisor in the research work also deserves acknowledgment for his cooperation and support. I found him encouraging and accommodating, in the meetings.

I am also thankful to Dr. Kevin K.W. Ho, co-editor (-in-chief) of the Library Hi Tech Journal, for his constructive and prompt reviews of the manuscript of my first research effort. Dr. John P. A. Ioannidis and his co-researchers also deserve acknowledgment, for sharing the researchers' data (a derivative of the Scopus data repository labeled as the World's top two percent researchers) for public use.

The external evaluators – Prof. Dr. M. Emre Celebe (University of Central Arkansas, USA.) and Dr. Yi Bu (Peking University, China.), of this research effort, also deserver my gratitude as they constructively reviewed the document and helped in its refinement.

The meticulous review of the final manuscript by Dr. Hikmat Ullah Khan (Comsats University Islamabad), and Dr. Arif Ur Rehman (Bahria University Islamabad) should also be acknowledged, as it refined the document further.

It is also necessary to remember the kind responses of Dr. Abd-ul-Mateen and Dr. Muhammad Shiraz, at the Federal Urdu University of Arts, Sciences, and Technology, Islamabad, for offering me the lectureship opportunity, during times of financial difficulty.

My friends and colleagues – Malik Khizar Hayat, and Abd-ur-Rehman Janjua also deserve acknowledgment, who not only encouraged me in the difficult times but provided technical support, in my hour of need.

Finally, the management of International Islamic University Islamabad deserves recognition for offering scholarships to its employees and their descendants. About half of my academic career was realized because of the generous scholarships offered by the university.

In the nutshell, it was the blessings of Allah Subhanahu-Wa-Tahala, my parent's prayers, the guidance of my supervisors, and the support and, encouragement of my family, well-wishers, and friends that enabled me to complete this piece of work successfully.

Muhammad Sajid Qureshi

November 2022

Declaration

"I hereby declare that this thesis, neither as a whole nor as a part thereof, has been copied out from any source. It is further declared that no portion of the work presented in this report has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning."

Muhammad Sajid Qureshi

Dedicated to the people,

who contributed to my education.

Abstract

In the last fifteen years, the process of ranking Higher Education Institutions has catalyzed the development of higher education, nationally and internationally. Starting with the holistic rankings of world universities in 2003, it has crossed the milestone of subject-specific rankings. Currently, various ranking systems are gauging and publishing the HEI rankings, including the three well-known international academic rankings – ARWU, THE, and QS. Nevertheless, academic rankings are facing criticism, because of various issues including the transparency and validity of ranking methodologies, limited coverage of HEIs, and linguistic and regional biases.

This research work is aimed at enhancing the creditability of the ranking process, through the fine-grained analysis of the objective ranking indicators extracted from publicly verifiable data sources. To achieve the objective, it proposed two ranking methodologies — *OpenRank*, and *ResRank*. These methodologies generate HEIs rankings by employing the objective indicators extracted from publicly verifiable data sources including *Scopus, Google Scholar Citations, ArnetMiner,* and *DBpedia* data repositories. The research work also demonstrated the effectiveness of the fine-grained analysis of the ranking indicators, by producing the *subject-specific* rankings of Global HEIs in four scientific disciplines. The analysis highlights the position of the top 100 global HEIs for their *research productivity*, *research impact,* and *research contribution* of the researchers affiliated with them. The *sub-discipline-specific* rankings in the Computing discipline are also a significant outcome of the analysis.

Evaluation of the proposed *OpenRank* methodology advocates its effectiveness, transparency, and quick reproducibility. The *fine-grained analysis* approach makes the ranking results more representative. The *sub-discipline-specific* rankings are more decision-supportive as they expand the limited picture of the global HEIs and provide a better insight into their performance in various subdisciplines of Computing.

Keywords: Entity Ranking; Academic Rankings; Subject Specific Rankings of HEIs; Sub-Discipline-Specific Rankings; Bibliometrics; Scopus; Google Scholar Citations; Linked Data.

Contents

Achievements and Contributions

Journal Publications

1. **M. S. Qureshi**, Ali Daud, Malik Khizar Hayat *"Heading Towards the Sub-Discipline Specific HEIs Rankings."* ***IEEE Transactions on Computational Social Systems***, **2022.Pages**: 1-20 (early access pagination). **ISSN**: 2329-924X. **DOI**: https://doi.org/10.1109/TCSS.2022.3206396

2. **M. S. Qureshi**, Ali Daud, *"Fine-Grained Academic Rankings: Mapping Affiliation of the Influential Researchers with the Top Ranked HEIs."* **Scientiometrics, 2021, Vol.** 126, **Pages**: 8331–8361. **ISSN**: 1588-2861,0138-9130. **DOI**: https://doi.org/10.1007/s11192-021-04138-z

3. **M. S. Qureshi**, Ali Daud, Malik Khizar Hayat, Muhammad Tanvir Afzal, "*OpenRank - A Novel Approach to Rank Universities Using Objective and Publicly Verifiable Data Sources.*", **Library Hi-Tech journal, 2021, Vol:** 39(1). **ISSN**: 0737-8831. **DOI**: https://doi.org/10.1108/LHT-07-2019-0131

Under Compilation

4. **M. S. Qureshi**, Ali Daud, *"Academic Rankings of HEIs – A Survey of the Past and Future Trends."*, **2022**.

List of Tables

List of Figures

List of Abbreviations

APP	A metric that considers the contribution of the author(s) in a research publication
ARWU	Academic Ranking of World Universities
CC	The count of research citations obtained by the publications of a researcher
DBpedia	The data repository maintained by Wikipedia
GAR Score	The average of the rank scores allocated by the three academic rankings systems to an HEI, in the stipulated years.
GSC	Google Scholars' Citation repository
HEIs	Higher Education Institutions
H-Index	The metric invented by Jorge Hirsch in 2005 jointly measures the research productivity and research impact of a researcher.
HITS	Hyperlink-Induced Topic Search
IC	Information Content
League Table	The table shows ranking based on some well-defined ranking methodology, of institutions like universities, colleges, hospitals, etc.
LOD	Linked Open Data or Linked Data
OpenRank	The proposed academic ranking methodology is based on objective ranking parameters extracted from public data sources.
OWL	Web Ontology Language
PC	The count of research publications by a researcher
QS	Quacquarelli Symonds
RDF	Resource Description Framework
ResRank	The proposed methodology for rank the research scholars based on their *Publication Count, Citation Count, H-Index* score, and count of *Authors Per Publication.*
Scopus	The bibliographic data repository maintained by Elsevier
THE	Times Higher Education
URAP	University Ranking by Academic Performance
WT2P	The list of the World's top two percent researchers was published by a researcher at Stanford University, USA.

Chapter 1

Introduction

Educational institutions, especially those that deal with higher education, are considered among the catalysts of the national and global development process. Governments of developed countries nurture research, innovation, and commercialization through the HEIs. Moreover, they also have adopted systematic ways to gauge and evaluate the academic performance of their HEIs. 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 Ranking from the Netherlands, University Ranking by Academic Performance (URAP) from Turkey, and the Jeddah Ranking from Saudi Arabia. Usually, these evaluation efforts declare their results in terms of the HEIs ranking scores. The first ranking of the global HEIs was produced by ARWU in 2003 [1].

The ranking process has gained substantial acceptance among the major stakeholders [1, 2]. It has changed settings of the higher education and is likely to cause further development, nationally and internationally. Being a catalyst for global competition, rankings have gained attention in major countries of the world. Now, the rank of an HEI is assumed a vital image-builder of an HEI, in the academic sector.

The interest of academic experts in the rankings process also has increased in the last ten years. 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, academic experts published fewer than twenty articles on the topic, in research journals; while in the year 2019, they published more than 100 articles, according to the Scopus database [2].

1.1. The Academic Performance Evaluation Process

The process of academic ranking usually involves a complex ranking methodology that is realized in terms of ranking indicators. The ranking indicators reflect the aims and objectives of the ranking entity.

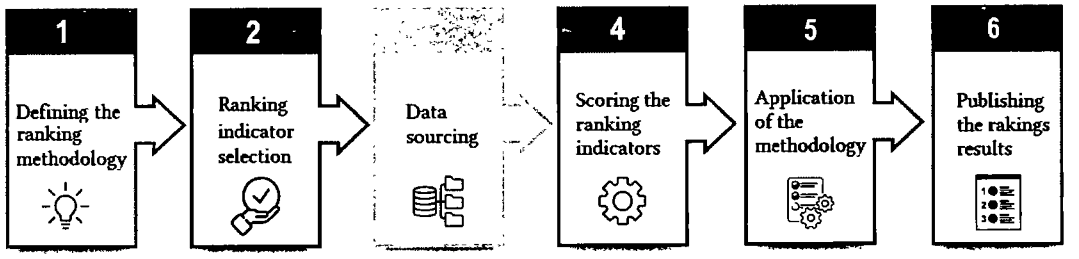


Figure 1.1 Steps of the Academic Ranking Process

The indicators determine the effectiveness of a ranking system and differentiate it from other methodologies. An indicator represents and gauges a particular dimension of a ranking methodology. Academic ranking 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 [3]. An effective academic ranking methodology assists potential students and faculty in choosing an HEI that is outstanding in a particular scientific discipline. The methodology also helps the academic managers in the better management of their areas of interest and provides decision-supporting analytics to the sponsors of the HEIs [24]. An effective academic ranking methodology employs comprehensive ranking indicators and prefers quantitative measures over qualitative measures. It exploits publicly verifiable data and provides a fine-grained analysis to the users. Scopus, DBpedia, GSC, and ArnetMiner repositories are examples of publicly verifiable data sources.

1.2. The Rationale for the Academic Rankings

Performance evaluation is essential for all businesses and productive organizations, irrespective of the domain and nature of the activity. Major stakeholders including the business owners, technical professionals, business regulatory authorities, etc. develop various regulatory authorities and quality management systems, to define, evaluate and control the desired quality. These systems help to maintain and improve the quality standard, hence performance and productivity.

The academic world is also developing and implementing various performance and quality control systems. These systems have a vital role in the academic improvement of any academic unit, irrespective of the size, such as an academician, department, faculty, or institution. Moving the scale further, modern academic performance evaluation systems are gauging the national, regional, and global performance of HEIs [4]. Being a catalyst for the growth of the national economies, the higher education sector is vibrant globally. The ranking systems have become an effective tool to gauge the performance of the education sector. Usually, the ranking systems measure research productivity and teaching quality. Some other rationales for the ranking process are listed below.

i) Academic performance evaluation helps the HEIs in comparing their performance with those of similar institutions at local, regional, and global levels. Identifying their strengths and weaknesses requires them to decide the important factors, in determining academic quality and research productivity. The comparison helps them, in leveraging their comparative advantage in the education sector [5].

ii) A better ranking score of an HEI advocates its performance, thereby helping it, win performance-based funding, accreditation reviews, academic program reviews, and national benchmarking process.

iii) Academic managers employ the ranking results in the academic policy and decision-making process.

iv) The subject-specific academic rankings substantially support the university-major selection by the prospective students, employers, and policymakers of the professional training programs [6].

v) Academic rankings also indicate the HEIs' performance to the Non-Governmental Organizations (NGOs) interested in educational funding.

1.3. The Need for Objective and Transparent Rankings

Academic rankings have gained overwhelming popularity in academia and their influence is yet increasing. Nevertheless, academic experts have shown considerable controversies regarding the validity and transparency of the rankings. They pointed out various debatable issues of the reputed rankings including the following:

- Controversial ranking methodologies [7, 3]
- Use of the "subjective" ranking indicators [7].
- Validity and transparency of the ranking results [5]

- Misinterpretation of academic performance [8, 9]
- Limited coverage of the global HEIs [10]

A sample of HEIs' rankings in the year 2017, having significant differences, by the three reputed international academic ranking systems is presented in chapter 3 (please consult section 3.2). The significant differences in ranking results of various ranking entities support reservations of the academic experts.

1.4. Open Research Issues

In Section 1.7, some well-known academic ranking systems were discussed. Table 3.1 also highlighted the significant differences among the results of their rankings. Academic experts have highlighted various issues of the ranking process, as in [5, 7, 8, 4]. Here is a brief description of the research issues.

a) Controversial Ranking Criteria

Usually, the HEIs ranking lists are based on weighted combinations of scores generated by academic performance evaluation indicators. Such a holistic single-valued score hides many aspects of the performance of an HEI. Moreover, setting the weights of the indicators is controversial [8, 4]. The unfair comparisons of HEIs having very different sizes, budgets, locations, etc. also weaken the validity of the ranking process. Similarly, the nomenclature issues of HEIs and the availability of their data required data, make the ranking indicators and their weights a debatable issue.

b) Validity and Transparency of the ranking results

Research output and quality should be assessed using verifiable and quantifiable criteria. Academic rankings being a transparency tool, are expected to be highly transparent and comprehensive as usually required to research publications. Nevertheless, certain academic rankings are not transparent enough in explaining the calculation of the holistic final score obtained by an HEI [8].

c) Limited coverage of the HEIs

The current academic rankings do not cover the diversity of higher education, and, in practice, their lists contain about 5-8 percent of HEIs worldwide. Originally, the rankings were intended to highlight the performance of global HEIs; therefore, they have intrinsic limitations in accurately representing the regional and national HEIs. Institutions having a better presence on the Internet get more favor. Similarly, some disciplines such as Engineering, and Computer Science are getting better representation (resulting in a drawback for generalist universities over applied ones) [4]. The ARWU academic ranking system employs ranking indicators like "*Alumni as Nobel laureates & Fields Medalists*" and "*The staff as Nobel Laureates & Fields Medalists*". Such outstanding contributions are rare and not applicable to many global HEIs. These parameters exclude many HEIs with outstanding academic performance, from the list of top-scoring institutions.

d) Unavailability of the data

The scale of a ranking project also affects the feasibility of a ranking methodology. In a large-scale project, comparing the performances of universities worldwide, difficulties lie in accessing certain university data and ensuring the validity of the comparisons. Not all university data is publicly available, and some universities may reject requests of the evaluator, for the required information [10].

e) Misinterpretation of the academic performance

Even when data are available, comparing university performances of different countries or regions can be problematic. University performance is affected by the larger socio-cultural and politico-economic context. For example, reputation-based evaluation can be highly biased toward famous universities or universities in reviewers' home countries. Another major criticism faced by the rankings is that some indicator names do not generally refer to what they measure. For example, the ARWU ranking methodology employs a ranking indicator "*research quality*" by only measuring some research results. Furthermore, when referring to research excellence regarding publications, the selection of publications in just Science and Nature raises controversy. Similarly, the justification of the indicator's weighting is debatable [5, 7].

f) Subjective Ranking Indicators

Currently, rankings methodologies employ a combination of "objective indicators" (i.e., staff, library, budget) and "subjective indicators" (opinions on institutional prestige, Academic Peer Review, Employer Reputation, etc.). The use of subjective indicators significantly affects the final score of an HEI [3].

Table 1.1 Baseline Academic Ranking Methodologies

	Ranking System	Publisher	First Ranking
1	Academic Ranking of World Universities	Shanghai Jiao Tong University	2003
2	THE University Ranking	Times Higher Education	2007
3	QS World University Rankings	Quacquarelli Symonds Limited	2004

Academic ranking is not a new phenomenon, as various international academic ranking systems are gauging the performance of the HEIs and publishing their results regularly. The ranking methodologies of the following well-known academic ranking systems shall be used as the baseline measures, to compare the results of the proposed ranking methodology.

The listed benchmark academic ranking systems are widely used and well-known in the academic world. For a reliable evaluation of the proposed methodology, its results shall be compared with those, produced by the baseline ranking systems. The benchmark ranking systems have been discussed in the first part of the literature review (Section 2). Detail of their ranking methodologies and indicators can be consulted there.

1.5. The Research Problem

After a discussion on the ranking methodologies of some well-known academic ranking systems and pointing out various issues of the ranking process, the following research questions are posed.

1. To what extent, the less controversial academic rankings of global HEIs would be produced, using the objective parameters extracted from publicly verifiable data repositories?
2. Is it feasible employing publicly verifiable data sources to increase coverage of the global HEIs, in the ranking process?
3. Does the fine-grained analysis of academic ranking parameters provide more valid and representative academic rankings of global HEIs?
4. Can the existing electronic repositories be exploited for producing the sub-discipline-specific rankings of global HEIs?

Although the listed research issues are recognized by academic experts in the literature. Nevertheless, information technology professionals having sound technical knowledge and familiarity with the academic ranking domain, are in a better position to provide a functional solution.

1.6. Proposed Solution

To achieve the desired research objectives, the research analysis is divided into three modules. Each module or part is described in detail, in the upcoming chapters after the description of the related work. Here is a brief introduction to the modules.

a) Ranking HEIs Using Objective & Publicly Verifiable Data Sources

As per the research objectives, this research module was intended to enhance the creditability of the ranking process, by using quantitative indicators based on publicly verifiable data sources. A new ranking methodology— OpenRank was developed, which ranks the global HEIs using the objective indicators extracted from two well-known publicly verifiable data repositories — the ArnetMiner and DBpedia. The resultant academic ranking reflects common tendencies of the international academic rankings published by THE, QS, and SRC. Evaluation of the proposed methodology advocates its effectiveness, transparency, and quick reproducibility. A comprehensive description of the research module, its objectives, contributions, and research methodology, is provided in Chapter 3. The chapter also presents the results and analysis and suggests practical implications.

b) Heading Towards the Fine-Grained Academic Rankings

This research module focuses on enhancing the credibility of the ranking process through a more fine-grained analysis of the academic ranking indicators. The proposed fine-grained analysis drives the objective indicators from two well-known publicly verifiable data repositories — Google Scholar Citations and DBpedia. The resultant academic rankings w.r.t the *Research Faculty, Research Productivity*, and *Research Impact* make the ranking process more transparent and fine-grained. The analysis also helps in understanding the areas of interest focused on by the reputed international academic rankings by ARUW, THE, and QS. The results of the fine-grained analysis advocate its effectiveness, in terms of transparency and validity. The research work is reported completely in Chapter 4, with a description of the research objectives and the research contributions. The research methodology and the data sources employed in the analysis are also discussed there. The chapter also presents the results and discussion along with the conclusions and future implications.

c) Heading Towards the Sub-Discipline Specific Rankings

This research module focused on the fine-grained analysis of the academic rankings in the Computing discipline. It analyzed the subject-specific rankings for the *research productivity*, *research impact,* and *research contribution* of the researchers affiliated with the top-ranked HEIs. The fine-grained analysis employs the highly curated data produced and published by the three well-known international academic rankings – ARWU, THE, and QS in the years 2018, 2019, and 2020. The researchers' profiles are obtained from the Scopus repository, while the DBpedia repository is employed for the information about HEIs and countries. For a stable comparison of the subject-specific rankings of global HEIs, by the three ranking systems, the Grand Average Rank (GAR) was employed. Whereas for finding the influential computing researcher the *ResRank* measure is used in the analysis. The sub-discipline-specific academic rankings provide a better insight into the academic performance of the global HEIs in various subdisciplines of Computing. So, they would be more decision-supportive due to the expansion in the limited picture of the global HEIs. The analysis focused on sub-discipline-specific rankings is among the first few such efforts. A comprehensive description of the research module, its objectives, contributions, and research methodology, is provided in Chapter 5. The chapter also presents the results and analysis and suggests practical implications.

1.7. The Research Contributions

Keeping in mind the research problem, the research works were divided into three modules that are reported in detail, in their respective chapters. The three research modules have multiple research contributions. Some of the contributions are common in the research modules as they share common research objectives. For a quick overview, the research contributions are listed here, in brief.

a) Ranking the Global HEIs Using Objective & Publicly Verifiable Data Sources

This research module was intended to make the following contributions:

- Enhancing the creditability of the ranking results by employing the quantitative performance indicators extracted from publicly verifiable data sources.
- Developing an academic ranking methodology based on quantitative indicators using two well-known data repositories, DBpedia and ArnetMiner.
- Demonstrating the effectiveness of the methodology on real data sources.

b) Heading Towards the Fine-Grained Academic Rankings

The research contributions of the second module, are summarized below:

- A fine-grained analysis of the academic data w.r.t the *research-productivity*, and *research-impact* of the HEIs, using publicly verifiable data sources.
- Producing the subject-specific rankings of the global HEIs, 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 and Google Scholar's Citations.

c) Heading Towards the Sub-Discipline Specific Rankings

This research module was intended to make the following contributions:

- Enhancing the creditability of the academic ranking process, by fine-grained analysis of the *research productivity*, and *research impact* of the HEIs, using publicly verifiable data sources.
- Producing the subject-specific rankings of HEIs, 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 and Google Scholar's Citations.

1.8. Some Well-Known Global HEIs Ranking Systems – A Review

Rankings are not a new phenomenon, although their recent proliferation is, especially in the domain of HEIs. Historically, the credit for publishing the first academic rankings is given to the *US News and World Report* magazine, which published the "*American Colleges Ranking*" in 1983 [7]. It was the first institution-based effort, to evaluate and rank academic institutions. Nevertheless, the first ranking of the global HEIs was produced by ARWU in 2003 [1]. Today, dozens of national and global academic ranking systems have emerged, been established, and recognized. After 2006, ARWU used the following academic areas, while assessing hundreds of HEIs worldwide.

Table 1.2 provides a list of some well-known academic ranking systems, their publishers, and the base countries. The ranking systems have diverse criteria or ranking methodologies to measure the performance of the HEIs. Having an in-depth review of all the ranking systems, along with their ranking methodologies, is beyond scope of this research work. Nevertheless, a brief but comprehensive introduction to some well-known global HEI ranking systems is following.

a) Academic Rankings of World Universities (ARWU)

ARWU Rankings (or Shanghai Rankings) is an annual publication of HEIs rankings by Shanghai Ranking Consultancy, based in China. The Chinese government-funded academic ranking system was initially developed to measure the gap between Chinese and the "world-class" universities. Today, ARWU rankings are comparable with the *QS World University Rankings, Leiden Rankings,* and *Times Higher Education World University Rankings*. These academic rankings are widely observed and considered influential internationally. The first ARWU global academic ranking, having diverse indicators, was compiled, and issued by Shanghai Jiao Tong University (SJTU), in June 2003. After 2006, ARWU used the following academic areas, while assessing hundreds of HEIs worldwide.

Table 1.2 Some Well-Known Global Academic Ranking Systems

	Ranking System	Publisher	Country	1st Ranking Year
1	Academic Ranking of World Universities (Shanghai Ranking)	Shanghai Jiao Tong University	China	2003
2	QS World University Rankings	Quacquarelli Symonds Ltd.	England	2004
3	Leiden Ranking	Leiden University, Netherland	Netherland	2007
4	Webometrics Ranking of World Universities	Cybermetrics Research Group	Spain	2009
5	SCImago Institutions Rankings	SCImago Research Group	Spain	2009
6	THE World University Rankings	Times Higher Education	England	2010
7	University Ranking by Academic Performance (URAP)	Middle East University of Technology (METU), Turkey	Turkey	2011
8	Round University Rankings	RUR Ranking Agency	Russia	2010
9	CWUR University Rankings	Center for World University Rankings, Jeddah	UAE	2012
10	4icu.org University Web Ranking	4 International Colleges &Universities	Australia	2005

In 2009, ARWU ranked a hundred HEIs, for the subject categories, including Mathematics, Computing, Business, Physics, and Chemistry. Currently, the ARWU is following the ranking methodology [11], based on the indicators listed in Table 1.2.. The ranking system has an objective ranking methodology, although it pays more attention to research productivity in the scientific disciplines. Social science disciplines, academia-industry interaction, and teaching quality get less coverage in the ARWU methodology.

b) The Quacquarelli Symonds (QS) World University Rankings

The QS academic rankings is an annual publication of university rankings by Quacquarelli Symonds Limited. Before 2009, it was known as THE-QS (Times Higher Education and Quacquarelli Symonds) rankings, as it collaborated with Times Higher Education (THE) magazine to produce the rankings from 2004 to 2009. THE is a United Kingdom-based weekly magazine that covers issues related to higher education. QS is an independent international organization focused on academia. It is the only influential worldwide academic ranking, published by a private company [10].

In 2009, QS published the academic ranking as the *QS World University Ranking*. Now, it covers the global-overall and subject-specific rankings, alongside the regional ranking lists for Asia, Latin America, and the BRICS countries (Brazil, Russia, India, China, and South Africa) with explicitly defined ranking methodologies. The QS ranking methodology groups its ranking indicators into four categories – *Research Quality, Teaching Quality, Graduate Employability*, and *International Outlook.*

The QS rankings published in 2009, covered the Arts & Humanities, Life Sciences & Biomedicine, Natural sciences, social sciences, and Technology. The rankings contained 600 global HEIs according to the final scores and assessed 300 HEIs for their subject categories. Currently, the QS is employing the ranking methodology [12] based on the indicators listed in Table 1.3. The QS World University Rankings are among the most widely read university rankings; Nevertheless, it faces criticism for giving undue weight to subjective indicators like *'Academic Peer Review'* and *'Employer Reputation' etc.* Being an independent organization, it is also blamed as commercially influenced [7].

c) Times Higher Education (THE) Academic Rankings

The Times Higher Education magazine has been publishing the academic rankings under the title, *THE World University Rankings* on annual basis, since 2009. Earlier, the magazine had a joint venture with the QS corporation to produce THE–QS World University Rankings from 2004 to 2009. THE current ranking methodology employs the well-known academic repositories *Thomson Reuters* and *Elsevier.*

The magazine annually publishes, the world-overall, subject-specific, and reputation-based rankings along with the regional league tables for Asia, BRICS countries, and Emerging Economies. The ranking announced in 2010-2011, employed the following ranking indicators clustered under five categories – Teaching, Research, Citations (research impact), International Mix, and Industry Income. Table 1.4 summarizes the ranking indicators [13].

Table 1.3 ARWU Ranking Methodology

Criterion	Indicator	Code	Weight	Data Source
Quality of Education	Alumni as Nobel laureates & Fields Medalists	– Alumni	– 10%	official websites of Nobel Laureates & Fields Medalists
Quality of Faculty	The staff as Nobel Laureates & Fields Medalists Highly cited researchers in 21 broad subject categories	– Award – HiCi	– 20% – 20%	official websites of Nobel Laureates & Fields Medalists Thomson Reuters' survey of highly cited researchers
Research Output	Papers published in Nature and Science* Papers indexed in Science Citation Index-expanded and Social Science Citation Index	– N&S – PUB	– 20% – 20%	Citation index
Per Capita Performance	Per capita academic performance of an institution	– PCP	– 10%	National Agencies such as the National Ministry of Education, National Rector's Conference, etc.

* Not applicable to institutions specialized in humanities and social sciences whose N&S scores are relocated to other indicators.

Table 1.4 QS Ranking Methodology

Indicator	Weight	Elaboration
Academic Peer Review	40%	Based on an internal global academic survey
Faculty/Student Ratio	20%	A measurement of teaching commitment
Citations Per Faculty	20%	A measurement of research impact
Employer Reputation	10%	Based on a survey of graduate employers
International Student Ratio	5%	A measurement of the diversity of the student community
International Staff Ratio	5%	A measurement of the diversity of the academic staff

THE academic rankings are widely observed and enjoy international recognition. It has a vigorous academic ranking methodology, inclining science-based institutions with relatively few undergraduates. HEIs that have more comprehensive study programs and undergraduates are less favored by the methodology. THE ranking methodology faces criticism for favoring the English HEIs and for the commercialization of the ranking process [7, 8]. Overall, it is an improved version of the QS raking methodology.

Table 1.5 THE Ranking Methodology

Indicator	Elaboration	Weight (%age)
Industry Income – Innovation	- Research income from industry (per academic staff)	- 2.5
International Diversity	- The ratio of international to domestic staff - The ratio of international to domestic students	- 3 - 2
Teaching – The Learning Environment	- Reputational survey (teaching) - PhDs awards per academic - Undergrad. admitted per academic. - Income per academic - PhDs/undergraduate degrees awarded	- 15 - 6 - 4.5 - 2.25 - 2.25
Research – Volume, Income, and Reputation	- Reputational survey (research) - Research income (scaled) - Papers per research and academic staff - Public research income/ total research income	- 19.5 - 5.25 - 4.5 - 0.75
Citations – Research Influence	- Citation impact (normalized average citation per paper)	- 32.5

d) The Academic Ranking Methodologies and Indicators

The academic ranking of HEIs is a complex process that involves six steps [4] as illustrated in Figure **1.1**. The process employs a ranking methodology that is realized in terms of ranking indicators that reflect the aims and objectives of the ranking entity [10]. An effective ranking methodology:

a) Employs comprehensive parameters
b) Prefers quantitative measures over qualitative measures
c) Uses the publicly verifiable data sources
d) Enables the users to "Drill Down and Roll Up" the indicator's detail.

Presently, there exist dozens of national (regional) and international academic ranking systems. Their diversity itself is evident in the fact that none of them has global acceptability, because of their limitations. The ranking systems use various ranking dimensions. For example, the *Overall Performance, Research Productivity, Education Quality, Average Performance,* and *Internationalization Factor*. Some ranking systems dynamically rank the HEIs, based on user-defined criteria [4, 7]. Nevertheless, a separate

Table 1.6 Major Categories of the Academic Ranking Indicators

Category	Indicator	Description
Research Activities	Publication Count (Research articles)	Research Productivity
	Publication Per faculty (Per Capita Output)	Scientific Productivity
	Highly Cited Publications	Research Quality
	Citations Count	Research Quality
	Citation Per Faculty	Current Scientific Productivity
	Total Scientific Publications	Scientific Productivity
	Total Publications Per Faculty	Faculty Research Productivity
	Patent's Value	Invention & Innovation Effort Assessment
Education Quality	Faculty Count	Number of permanent faculty members
	Student-Faculty Ratio	Assessing Faculty Per Student
	Academic Facilities	Research Laboratories, Technology Incubation Centers, etc.
	Ph.D. Faculty	The ratio of the Ph.D. faculty and full-time faculty
	Highly Cited Researchers	Count of the renowned research scholars
	Faculty H-Index	Average H-Index of researchers in an HEI
	Nobel Laureates	Nobel prize winner research scholars
Student Body	Total Students	Number of the students currently enrolled
	Ph.D. Graduates (Count or % age)	The ratio of the Ph.D. students to the total students
	Award Holders & Entrepreneurs	Award Holders & Entrepreneurs Alumni
	Graduate Employability	Assessment of acceptability of graduates
	Alumni Strength	Count of number of graduates of an HEI
Public Opinion	Peer Institutional Review	Assessment of performance of an HEI by another HEI of the same domain and area
	Academic Expert Opinion	Assessment of academic experts about an HEI
	Student's Opinion	Assessment of the student's judgments
	Student Attraction	Assessing the student's favorite HEIs
	Employer's Opinion	The reputation of an HEIs among the employers
Financial Strength	Annual Budget	Assessment of available financial resources
	Available R&D Grants	Assessment of available financial resources for the research and development activities
	Endowment Funds	Assessment of reticent sources of income for an HEI
International Factor	International Faculty	The ratio of the faculty of international repute to the total permanent faculty of an HEI
	International Students	The ratio of the international students to the total full-time (regular) students of an HEI
	International Repute	Assessment of international repute based on various HEIs

discussion on the ranking dimensions and indicators of every ranking system is beyond scope of this research effort. Commonly used ranking dimensions and ranking indicators are listed in Table 1.6. The ranking results are represented for a specific cumulative measure and, then by different categories like the following [9]:

a) Based on the commonly available indicators
b) Cumulative points, global rank, top 100, etc.
c) Categorical score based on specific indicators – Subject-specific rankings, regional comparison.
d) Available budget & research grants, endowment funds
e) Resourcefulness i.e., faculty quality and size, research laboratories, campus size

1.9. Related Data Sources

A major criticism faced by existing academic ranking methodologies is the use of data sources that are not publicly verifiable [7, 14]. As there is no purposely-built public data repository that would serve the intended academic ranking methodology. Consequently, it was decided to employ different data sources that stratify the necessary conditions. A brief description of the data sources is given below.

a) The DBpedia Dataset in Linked Data Cloud

DBpedia is a structured data repository maintained according to the Resource Description Framework (RDF) semantic ontology. It is one of the largest constituents of the Linked Open Data (LOD) Cloud. The dataset is owned by Wikipedia – the well-known Internet-based electronic Encyclopedia that has become one of the fundamental knowledge sources of mankind. The Cloud enables access to many semantically described and digitally linked representations of objects and entities. The English version of the DBpedia describes 4.0 million "things", having 470 million affiliated "facts". These things are classified with a consistent ontology. This information can be used in improving existing data-hungry applications as well as in innovative data products.

The OpenRank methodology employed the DBpedia dataset[1] (using its SPARQL-Endpoint through SPARQL queries) to measure the values of the ranking indicators [15] for the ranking categories Academic Sustainability (AS) and International Factor (IF). Two sample SPARQL queries are given below to retrieve the HEIs' information available in the DBPedia. Other related queries (SPARQL queries[2] for DBpedia and SQL queries[3] for the ArnetMiner dataset) are uploaded to the repositories at GitHub[4].

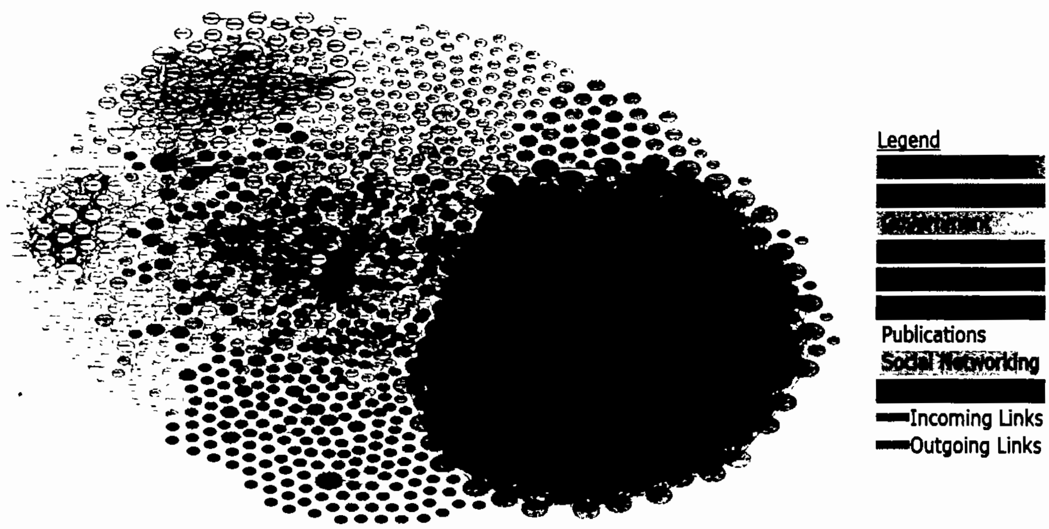


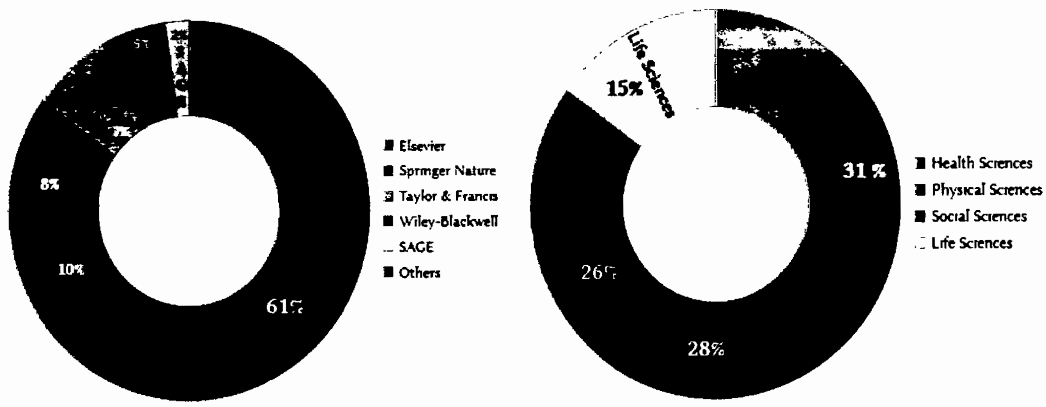
Figure 1.2 An Instance of the LOD Cloud

Sample SPARQL queries to retrieve the HEIs' information available in the DBpedia.

PREFIX dbr:<http://dbpedia.org/resource/>
PREFIX dbo:<http://dbpedia.org/ontology/>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT *
WHERE
{
dbr:Harvard_University
dbo:numberOfStudents ?TotalStudents;
foaf:homepage ?URL;
dbo:facultySize ?Faculty;
dbo:endowment ?Endowment;
dbo:numberOfUndergraduateStudents
?UGStudents;
dbo:numberOfPostgraduateStudents
?PGStudents.
}
	PREFIX dbr:<http://dbpedia.org/resource/>
	PREFIX dbo:<http://dbpedia.org/ontology/>
	PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
	
	SELECT *
	WHERE
	{
	dbr:Australia dbo:areaTotal ?Area;
	dbo:language ?Language ;
	dbo:populationCensus ?Population;
	dbo:currency ?Currency ;
	dbo:hdi ?HumanDevelopmentIndex ;
	dbp:gdpNominal ?GDP .
	}

A SPARQL data-retrieval query for extracting information related to HEIs from the DBpedia.	A SPARQL query for extracting information related to Countries from DBpedia.

1 https://wiki.dbpedia.org/services-resources/datasets/data-set-39
2 https://github.com/muhammadsajidqureshi82/ArnetMiner
3 https://github.com/muhammadsajidqureshi82/DBpedia
4 https://github com/muhammadsajidqureshi82/OpenRank



a) b)

Figure 1.3 Share of the a) Publishers and b) Serial titles indexed in the Scopus

b) The ArnetMiner data repository

The ArnetMiner dataset was created by Professor Jie Tang [16], from Tsinghua University China, in the year 2006, as a research project in social influence analysis in academia. The dataset was built to search and perform data mining operations on academic entities. For instance, the existence of connections between authors, research articles, publication venues, and research trends prevailing in the social networks of the researchers. ArnetMiner computing servers automatically extract the researchers' profiles and publications from online digital libraries using pre-defined heuristic rules. The citation data is extracted from different scholarly sources including DBLP and ACM.

ArnetMiner is considered a widely used and one of the best-curated databases for computer science articles [16]. The ACM-Citation-Network (version 8, covering data entries up to July 2016) includes 2,381,688 research articles, published by 1,712,433 researchers, and 10,476,564 citation relationships ranging from 1936 to 2016 [17]. The collaboration relationships constructed by the interaction of these authors, allow us to study the interaction, scientific collaboration, and influence among scientists. ArnetMiner has published several datasets for academic research purposes, including [17, 18, 19, 20].

c) Google Scholar Citation (GSC) data repository

GSC repository was created and is currently owned by Google incorporation. It was launched on November 20, 2004. GSC allows its users to search for academic resources whether online or in digital libraries. Google Scholar index includes most peer-reviewed online academic journals and books. It also maintains a record of conference papers, research theses, technical reports, and other scholarly literature, including court opinions and patents. Dominantly, GSC is about research publications representing papers published by research scholars.

d) Official Ranking Results by the Reputed Rankings Systems

As per the research objective, for fine-grained analysis of the academic ranking statistics, the existing ranking results by the three well-known ranking systems — ARWU[1], THE[2], and QS[3] were reused. These ranking results are publicly available on their official websites. Some of them also provide a limited amount of data in spreadsheets or flat files.

e) The Scopus Repository

The profiles of the influential researchers are extracted from the Scopus repository. Scopus is a reliable and comprehensive data source for large-scale analyses in research assessments. The Elsevier-led subscription-based data repository, is among the world's largest curated abstract and citation databases, with a wide global and regional coverage of scientific journals, conference proceedings, and books. Its data are indexed through rigorous content selection and re-evaluation. Along with the data about scientific articles, Scopus also maintains the author and institution profiles. Since its launch in the year 2004, the content of the database has grown to over 76 million records, covering publications from 1788–2019. Elsevier itself as a scientific publisher is a considerable contributor (about 10 percent) to the content of Scopus. Scopus provides the publication output, citation, and international collaboration data to various academic ranking systems including THE, Shanghai Ranking Consultancy (since 2015), and QS. The data is also employed by the Frankfurter Allgemeine Zeitung Economists Ranking in Germany and the National Institutional Ranking Framework (NIRF) in India [21] The statistics about Scopus summarized in Figure 1.3, are based on the work reported in [22]. The authors' data in the

---

1 http://www.shanghairanking.com/ARWU-Methodology-2020.html. [Accessed on Dec 2020]

2 https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2020-methodology

3 https://www.topuniversities.com/subject-rankings/methodology. [Accessed on Dec 2020]

Scopus repository are clustered into publication histories known as the Scopus Author-Profiles. Elsevier claims to have a service of about 12,000 credible authors to assess the quality of the data in Scopus. The author profiles are maintained to reflect the complete publication history of the researchers [23].

In our data analysis, the researchers' information was extracted from a derivative dataset of the Scopus repository. The dataset is published by a well-known researcher at Stanford University, United States [24] in the year 2020.

The rest of the document is organized in this way. A summary of the literature review is presented in chapter 2. The research activities and their outcome related to the newly developed academic ranking methodology OpenRank, are reported in chapter 3. The fine-grained analysis of the ranking data, subject-specific rankings HEIs in four disciplines, and the results of mapping affiliation of the influential researchers with the top-ranked HEIs are presented in chapter 4. The sub-discipline-specific rankings of the global HEIs in the computing discipline are reported in chapter 5. While chapter 6 concludes the thesis with a description of the research findings and their future implications.

Chapter 2

Literature Review

This chapter summarizes the research work related to the research thesis. For ease of understanding, the description is divided into three sections that represent the different dimensions analyzed and reported during this research effort. The first section describes the previous research efforts focusing on the ranking of the global HEIs using open data sources. The second section presents the research work focusing on the fine-grained analysis of the academic rankings. The third section describes the research efforts that analyzed subject-specific academic rankings by the three reputed rankings entities.

2.1. Ranking HEIs Using Objective and Publicly Verifiable Data Sources

This section describes the previous research efforts describing the ranking of the global HEIs using open data sources. Keeping in view the OpenRank methodology, the review of the related work is divided into two categories. The first category covers the conventional academic ranking systems i.e., ARWU, QS, and THE, academic rankings. While, the second category covers the academic ranking efforts, based on public data sources i.e., the academic ranking by "Webometrics" [25] and the use of Linked Open Data in ranking global HEIs [15]. It also overviews the recently proposed academic ranking methodologies by the knowledge workers and experts.

Most of the conventional academic ranking methodologies employ a combination of objective and subjective ranking indicators that reflect their predefined ranking objectives and scope. ARWU by SRC [11], *World University Rankings* by QS [12], THE *World University Rankings,* and the *University Ranking by Academic Performance* (URAP) by Middle East Technology University (METU) [13], belong to this category. Since our research methodology emphasizes the use of public data sources and quantitative parameters so the process of these ranking entities has limited relevance for us. Nevertheless, these methodologies collectively give a good idea of the ranking indicators and their significance or weights. Commonly used ranking indicators by these methodologies can be categorized as *Research Productivity or Quality, Academic Quality, Academic Strength or Sustainability*, and *International Factors*.

Usually, the conventional ranking methodologies employ data sources that are not publicly verifiable. Moreover, they use subjective ranking indicators like 'Academic Peer Review' and 'Employer Reputation' by QS and THE academic rankings. The use of controversial data sources and subjective indicators makes their rankings controversial [7]. When a ranking entity employs a ranking indicator that is either based on non-quantifiable opinion(s) or its value(s) is not publicly accessible, then the ranking results become controversial, especially for the competing HEIs.

G. Boulton in [10] and Mu-Hsuan Huang in [3], discussed the issue in detail. For example, the QS academic ranking methodology employs the qualitative indicator "*Academic Peer Review*" based on an internal global academic survey. Moreover, the methodology assigns a decisive weight (40 % of the overall score) to the indicator. A properly designed and carefully executed peer review would provide invaluable information. Nevertheless, the indicator faces criticism for its subjectivity. The definition of the "peer" and then its selection during the evaluation process involves substantial human interaction. The soundness of the judgments by the peers also faces criticism for its fairness.

Similarly, the methodologies assign significantly different weights to the ranking indicators. These differences affect the ranking results, and an HEI gets a different rank by different ranking entities even in the same year. A sample of differences among the reputed rankings for the year 2017, is presented in Table 1. Various stakeholders in the higher education sector have shown their concerns, about the strength of the connection between education quality and the ranking indicators employed by the ranking systems. These academic ranking methodologies are reviewed and analyzed by various experts, as in [5, 7, 8, 4].

Mu-Hsuan Huang especially focused on THE-QS and HEEACT and ARWU in his analysis of various academic ranking systems [3]. According to his findings, the results of different ranking systems vary, sometimes dramatically, due to their different methodologies and emphases of various ranking criteria. Subjective or qualitative parameters impressionistically favor certain universities and produce results drastically different from quantitative data-based rankings.

The second category of academic rankings bases its academic performance analysis, on web-based or public data sources and employs objective indicators. For instance, [26] ranked the global HEIs using the career outcomes of their graduates. They employed the

LinkedIn[4] repository containing the career information of millions of professionals across the world. Instead of combining multiple ranking dimensions in a single ranking statistic, the methodology focused on a single ranking criterion, exclusively. Although the dataset employed in the ranking methodology is not publicly accessible, nevertheless its data source – the LinkedIn data repository[5] qualifies for the condition of objectivity. The graduate's career outcome is of great significance in gauging the performance of an HEI, therefore incorporation of the ranking dimension is highly desirable in future academic rankings.

Rouzbeh M. et al. [15], ranked the HEIs using the structured data available in the Linked Data Cloud, through the Partial Information Content (PIC) measure. The Information Content (IC) based methodology ranks the HEIs using their PIC scores that are computed from the Linked data. Although the data source is publicly verifiable, nevertheless the PIC score is a crude measure to gauge HEI's performance, especially where the underlying data is not purposely curated.

Koen Frenken et al. [27] suggested that a holistic score computed by a benchmark system can be misleading in the academic decision-making process as the academic performance is influenced by the structural variables including HEI's size, age, location, disciplinary orientation, and country location. The authors explored three performance dimensions namely, *Research Excellence*, *Internationalization*, and *Innovation* using the data employed by CWTS Leiden Ranking. Based on their analysis, they argue that research performance differences among universities mainly stem from size, disciplinary orientation, and country location.

Therefore, HEIs' comparison is meaningful among universities of a similar size supplemented with contextual information. Their analysis also highlighted the need for

---

4 https://www.linkedin.com [Accessed in July 2020]

5 https://university.linkedin.com/

employing more fine-grained research benchmarking to obtain better support in the academic decision-making process.

The Webometrics ranking of world universities proposed by Aguillo et al. is another ranking that quantitatively analyses the Internet and Web contents related to the HEIs i.e., rich files (.doc, .pdf, .ps, etc.), external links, Web pages, and articles on Google Scholar repository [25]. Although the statistics employed by the methodology are peripheral, they can serve as a publicly verifiable, and objective ranking methodology.

Peter Haddaway et al. [28] proposed a new Global Research Benchmarking System (GBRS) for uncovering the fine-grained research excellence using quantitative and publicly verifiable academic performance indicators. According to the authors, the GBRS enables academic decision-makers to obtain sufficiently-processed benchmarking data showing variation among specific research disciplines and identify academic areas of strategic investment. Such fine-grained and publicly accessible data would enable meaningful insights into the performance of HEIs.

In [29], Yuxiao Dong et al. argued that over 90% of modern world-leading innovations are the result of perpetually increasing international research collaborations. It means, that HEIs with a substantial number of international faculty and post-graduate students are enhancing international research collaborations, thereby technology innovation. The argument is based on the exploration of the large-scale scholarly dataset sourced from Microsoft Academic Services – the Microsoft Academic Graph (MAG). Such public data sources would provide multiple key-performance indicators for the HEIs' ranking.

José Luis analyzed the collaboration behavior of associated researchers in research work and its impact on their research outcomes [30]. For this analysis, the author employed the emerging bibliographic source — Microsoft Academic Graph. According to him, researchers being part of sparse and thin networks have a higher research impact. Usually, these thin networks involve a limited number of isolated researchers with very effective collaboration. The analysis helps us infer the fact that the HEIs with better facilitation for dedicated researchers are producing high-impact research.

In [31] Arnab Sinha et al. employed the Microsoft Academic Search (MAS) dataset having at its core a heterogeneous entity graph comprised of six types of entities that model the scholarly activities: *field of study, author, institution, paper, venue*, and *event*. This data source is highly relevant to the intended research work. Like Google Scholar's Citations database, the academic graph contains abundant information about the researchers.

Table 2.1. Summary of the literature review for the research module I

Article Title	Author(s)	Venue	Year	Contribution (s)
What drives university research performance? An analysis using CWTS Leiden Ranking data	Koen Frenken et al.	*Journal of Informetrics*	2017	▪ Highlighted the limitations of the holistic ranking scores. ▪ The academic performance is influenced by structural variables including HEI's size, age, location, disciplinary orientation, and country location.
Uncovering fine-grained research excellence: The Global research Benchmarking System	Peter Haddway et al.	*Journal of Informetrics*	2017	▪ Proposed a new Global Research Benchmarking System (GBRS) for uncovering the fine-grained research excellence. ▪ They used quantitative and publicly verifiable academic performance indicators.
A Century of Science: Globalization of Scientific Collaborations, Citations, and Innovations	Yuxiao Dong et al.	KDD'17	2017	▪ The employed the Microsoft Academic Service data repository to extract the research collaboration among researchers of the HEIs.
Ranking Universities Using Linked Open Data	Rouzbeh M. et al.	LDOW'13	2013	▪ Ranked the HEIs using the Linked Data Cloud ▪ They employed the Partial Information Content (PIC) measure.
A comparison of three major academic rankings for world universities: From a research evaluation perspective	Mu-Hsuan Huang et al.	*Journal of Library & Information Studies*	2011	▪ They pointed out the use of subjective indicators and unusual weights for the indicators [6]. Like the QS methodology employs "Academic Peer Review" with a weight of 40 %

The work presented in [32] by Sheeja N.K. et al. emphasized the role of the scholarly output of an HEI in scoring rank in various ranking schemes. They analyzed the case of Indian HEIs with a high score in ranking based on the National Institutional Ranking Framework (NIRF) of India. The author employed the quantitative parameters extracted from publicly verifiable data sources including the official websites of NIRF, THE World University Rankings, and QS World University rankings. A summary of the literature review for the research module I is presented in Table 2.1.

Overall, the academic ranking efforts, covered in the second category, emphasize the need for a more fine-grain analysis of the key performance indicators of an HEI. Such a fine-grain analysis would better help the academic managers in the focusing area of their interest. At the same time, the analysis would support the stakeholders in the selection of the areas for strategic investment, which also offers them a competitive advantage.

2.2. Heading Towards the Fine-Grained Academic Rankings

This section presents the research work focusing on the fine-grained analysis of the academic rankings. Keeping in mind the research objectives – fine-grained analysis of the ranking parameters and use of the publicly verifiable data sources, the research efforts of two types are focused. First, that was aimed at the fine-grained analysis of the academic ranking dimensions. Second, the academic rankings efforts 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 [30]. According to the analysis, the dedicated researchers who are part of the sparse and small research networks, have higher research productivity and 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. [33], explored the relationship between leading scientists and top-performing organizations. They argued that up to 80% of the world's highly influential researchers work at 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 Abramo et. al. [34] 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, time-saving, and less costly.

The research work reported in [35] 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 are an extension of the research effort.

John Mingers et al. examined 130 HEIs in the UK by using Google Scholar Citations (GSC) data. They used citation-based statistics to produce the HEIs' rankings [36]. 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 [37], 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. This Context-sensitive analysis is another supportive voice for the proposed fine-grained analysis strategy.

The work presented in [18] by Sheeja N.K. et al. highlighted the role of Research Productivity in the academic ranking process. They studied the case of Indian HEIs which are highly ranked per the National Institutional Ranking Framework (NIRF) of India. According to their analysis, the most decisive performance indicator of an HEI is research productivity. The authors employed the quantitative parameters after extracting them from the official websites of NIRF, THE, and QS.

Jun Zhang et al. [38] 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* [27]. The work highlighted that these indicators are prone to conceptual ambiguity and uncertainty; nevertheless, in many cases, students and, HEI's managers consider them increasingly meaningful. Moreover, the difference in size, disciplinary orientation, and country location are major causes of the difference in 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.

Table 2.2. Summary of the literature review for the research module II

Article Title	Author(s)	Venue	Year	Contribution (s)
Mining the Web to approximate university rankings	Corren G. McCoy	*Information Discovery & Delivery*	2018	▪ Presented an alternative to the university ranking lists using the HEIs' data available on Twitter
Using Google Scholar institutional-level data to evaluate the quality of university research	John Mingers et al.	*Scientometrics*	2017	▪ Explored the Google Scholar Citations (GSC) data to evaluate 130 HEIs of the United Kingdom. ▪ They claimed credible results and overcame some of the obvious problems of the ranking produced by the UK Research Excellence Framework (REF) for the same HEIs
Ranking universities based on career outcomes of graduates	Navneet Kapur et. al.	*KDD '17,*	2016	▪ They used the LinkedIn repository for career profiles of the professionals ▪ Ranking of the global HEIs using the career outcomes of their graduates. ▪ Alternative to the holistic ranking scores, based on ranking-dimension-oriented analysis

Baris Uslu et. al. used the predictive approach of correlational research to drill down methodologies of the four academic ranking systems – ARWU, THE, QS, and URAP [39]. He employed the ranking results published by these systems in the year 2018, to specify an expanded set of prominent ranking indicators. In his opinion, these potential indicators cover the common aspects of the ranking systems and would help the university leaders to develop better strategies. The researcher also determined the percentage values of the indicators including Citation, Income, Internationalization, Prize, Publication, Reputation, and Ratios/Degrees. According to the findings of the analysis, Research Reputation contributes 73.71% to universities' ranking scores. While, amazingly, 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 the use of objective ranking parameters; therefore, it consolidates the philosophy behind our fine-grained analysis strategy. A summary of the literature review for the research module II is presented in Table 2.2.

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 HEIs. Such analysis would better help the various academic stakeholders in their decision-making process. The academic managers would focus on the area of their interest, considering the analysis.

2.3. Heading Towards the Sub-Discipline Specific Rankings

This section describes the research efforts that analyzed subject-specific academic rankings by the three reputed rankings entities. As per the research objectives – fine-grained analysis of the academic rankings and sub-discipline-specific academic rankings, two types of existing research efforts are focused on. First, that was aimed at the fine-grain analysis of the academic ranking dimensions or parameter(s). Second, the research efforts focus on the sub-discipline-specific academic rankings based on publicly verifiable data sources. A summary of the review is following.

Huang et al. performed detailed bibliographic comparisons between Web of Science, Scopus, and Microsoft Academic at the institutional level, to determine their robustness for ranking the global HEIs [21]. They found that the HEIs have significantly different data representations across the three data sources. Similarly, these data sources have considerable inclinations toward certain scientific disciplines. Such differences cause drastic changes in the rank positions of the HEIs. Therefore, the use of multiple bibliographic data sources is strongly recommended, in the evaluation of the HEIs. The findings highlight the need for publicly verifiable data sources and fine-grained analysis of academic data.

The work of Baris Uslu reported in [39], is highly relevant to the intended fine-grained analysis of the academic ranking data. The author examined the individual weights of the significant academic ranking indicators for guiding university leaders in developing better strategies. He highlighted the common aspects of international ranking systems by employing the ranking data of 224 HEIs, published in the year 2018, by ARWU, THE, QS, and URAP ranking systems. The researcher re-evaluated significant ranking indicators including citation, income, internationalization, prize, publication, reputation, and ratios/degrees. The analysis revealed an overwhelming significance of the *research reputation* with a contribution of 73.71% to universities' ranking scores. As per the findings, for the reputed rankings entities, the most important ranking dimensions are *research productivity, research impact,* and *academic reputation*. The author named the three ranking indicators collectively, the "research reputation" of an HEI. According to his re-evaluated HEIs' rankings, only 19 universities occupied the same position among the 224 universities. Based on the research analysis, his recommendations are worth consideration by the HEIs' policymakers and directors of the reputed academic ranking entities.

Michael McAleer et al. [40] explored the ranking data published by THE ranking system for HEIs in Japan for the years 2017 and 2018. This analysis was aimed at finding the relationship between the *size of the student body*, internationalization, and *ranking scores* of the HEIs. Based on their findings, the author reported the considerable statistical significance of the two ranking parameters. The analysis emphasized the need for context-sensitive rankings of the global HEIs, for their more representative rankings.

The research work reported in [35] 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 [41], 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 are an extension of the research effort.

Francesco A. M. et al. [42] highlighted the inconsistencies in ranking results of the well-known academic ranking systems. They proposed a new quantitative and publicly verifiable proxy measure to gauge the academic reputation of an HEI, using the PageRank algorithm. As the algorithms operate with hard data so it would be a sound and credible mechanism for recognizing the HEIs and their affiliated researchers. This is another effort for enhancing the credibility of the academic, by reconciling the qualitative evaluation of academic prestige using the research impact.

The LinkedIn repository contains the career information of millions of professionals across the world. Navneet Kapur et al. exploited the profiles of professionals from the LinkedIn repository [26], to assess the graduate employability of the global HEIs. While ranking the global HEIs, they focused on a single ranking dimension exclusively, in the analysis. The HEIs rankings, based on the career outcomes of their graduates would provide substantial proof of the quality of education at the institutions. Therefore, such fine-grained analysis is of great significance in gauging the performance of the HEIs. The research work is another supportive evidence highlighting the need for ranking-dimension-oriented academic rankings.

The work presented in [43, 44, 24] by PA Ioannidis et al. became relevant to our research as it employs the Scopus researchers' profiles, belonging to various scientific disciplines including Information Technology. According to their findings, the citation and other evaluation metrics lose their value when they are misused. Therefore, a more comprehensive and robust evaluation or ranking metrics is required. They ranked the researchers by applying a new metric – the "*composite indicator* "and published for the public a dataset of more than 150,000+ top scientists.

The dataset provides standardization on citations, h-index, co-authorship adjusted hm-index, the order of authors, and a *composite indicator.* According to them, the calculation method behind the composite indicator is robust and resistant to misuse. The authors also published science-wide author databases containing data about the authors who secured their positions among the "World's top 2-percent researchers". All these researchers are indexed in the Scopus data repository. Thanks to their data curation, as in the research analysis, various researchers' statistics were derived from the dataset. The information about the author's affiliation also helped us in mapping the affiliation of the influential researchers with the top-ranked HEIs in Information technology.

Table 2.3. Summary of the literature review for the research module III

Article Title	Author(s)	Venue	Year	Contribution (s)
Updated science-wide author databases of standardized citation indicators	PA Ioannidis et. al.	*PLoS Biol*	2020	▪ Publishing of a dataset of more than 150K World top 2-percent researchers.
A standardized citation metrics author database annotated for the scientific field.	PA Ioannidis et. al.	*PLoS Biol*	2019	▪ Use of a new comprehensive and robust evaluation metric – the "**Composite Indicator**". ▪ Exploited the Scopus repository for performance evaluation of the professionals
Size, Internationalization, and University Rankings: Evaluating and Predicting Times Higher Education (THE) Data for Japan	Michael McAleer et al.	*Sustainability*	2019	▪ Explored the impact of Size, Internationalization, on University Rankings. ▪ As per their findings, the HEIs rankings methodologies must consider the size and internationalization factor.
Do first rate scientists work at first rate organizations?	Bo Yang et al.	*Malaysian Journal of Library & Information Science*	2015	▪ Explored the relationship between leading scientists and top-performing organizations ▪ As per their findings, up to 80% of the world's highly influential researchers work at top-performing organizations.

Giovanni Abramo et. al. [34] 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 at Italian universities, and in all disciplines of the hard sciences. According to them, the indicator is more effective, time-saving, and less costly.

John Mingers et al. examined 130 HEIs in the UK by using Google Scholar Citations (GSC) data. They used citation-based statistics to produce the HEIs' rankings [36]. 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.

Giovanni Abramo et. al. [45, 46] explored the impact of the ratio of top scientists to academic staff on the competitive strength of the Italian HEIs. The author mapped the affiliation of the Italian top scientists in over 200 fields, with the Italian HEIs and ranked them by the ratio of top scientists to the overall faculty.

According to their findings, the association between the ranking lists by the ratio of top researchers and average productivity was strong, in nearly all disciplines of the Sciences. The HEIs having high rank by productivity also have a higher ratio of top scientists (TS ratios) to the faculty. Very few of the institutions deviated from the investigated trend.

Corren G. McCoy et al. offered an alternative to the university ranking lists published in U.S. News & World Report, THE, ARWU, and Money Magazine [47]. They employed publicly available HEIs' data from Twitter and academic sources to compute social media metrics to approximate conventional academic rankings of US universities. They used two measurements to rank 264 HEIs of the US — first, the *University Twitter Engagement* (UTE) score (the total number of primary and secondary followers affiliated with an HEI), and second, the *Endowment, Expenditures, and Enrollment* (EEE) score. The authors argued that these metrics could be viable proxies for ranking HEIs in the United States. These statistics could be reproduced when required, rather than relying on an annually generated ranking score. Moreover, they supported diversity in the ranking lists. Utilizing publicly available Twitter and academic sources data is another supportive voice for the proposed fine-grained analysis strategy.

The work presented in [32] by Sheeja N.K. et al. highlighted the research productivity role in computing the rank of an HEI. They studied the case of Indian HEIs which are highly ranked per the National Institutional Ranking Framework (NIRF) of India. The authors used quantitative parameters which are publicly available and extracted from the official websites of NIRF, THE, and QS.

To evaluate the HEIs' academic reputation, Francesco Alessandro et al. made use of a proxy measure and eventually, compared the findings with well-established impact indicators and academic rankings [48]. They analyzed the citation patterns among HEIs in five subject categories of Web of Science and applied the PageRank algorithm to the resulting citation networks. According to the authors, the reputation of references gives the pathway to scientific citations. Therefore, the PageRank algorithm would be used to analyze the citation networks reflecting the reputation of an HEI in a specific field. The PageRank, being a well-rooted algorithm, is a reliable choice for analyzing the patterns in citations among the research articles published in peer-reviewed venues. The analysis would help to recognize the reputation of the venues among the researchers and HEIs. Using publicly verifiable data, and employing quantitative parameters in ranking methodology, the proposed approach consolidates the philosophy behind the fine-grained analysis strategy. A summary of the literature review for the research module III is presented in Table 2.3.

The literature review would be summarized as the existing efforts covered in this section underscored the necessity for fine-grained analysis of the Key Performance Indicators of an HEI. A fine-grained analysis would make the academic ranking process more transparent and less controversial. Moreover, it would better help the academic stakeholders, in the focusing area of their interest. At the same time, the analysis would support the stakeholders in the selection of the areas for strategic investment, which would result in a competitive advantage.

2.4. Summary

The review of the related research work highlights the fact that more transparent and reliable ranking results are achievable by the use of objective and publicly verifiable rankings indicators. Whereas giving substantial weight to subjective indicators leave more space for favoritism.

The fine-grained analysis makes the ranking results more representative and the ranking process less controversial. The holistic ranking score, even in the subject-specific rankings, hides many valuable aspects of a research institution. There are other ranking dimensions like the affiliation of *influential research faculty* with an HEI, and *graduate employability*, that can be explored for a better understanding of the educational dynamics.

Various HEIs perform remarkably well at the sub-discipline level, although they are not present on the top of the lists produced by well-known academic ranking entities. The sub-discipline-specific academic rankings are the need of the hour.

Context-aware academic rankings would be more representative while overlooking the context would result in biased judgments. Considering the contextual factors including the economic condition of the base county, academic age, languages, and size significantly affect the position of an HEI in the ranking lists.

The use of diverse data sources facilitates a comprehensive understanding of the factors that catalyze higher education. It is worth mentioning that although, the academic data available on the Internet is abundant and perpetually increasing. Nevertheless, most of the data repositories are not purposely built for academic ranking.

Chapter 3

Ranking the Global HEIs Using Objective and Publicly Verifiable Data Sources

In the last fifteen years, the process of ranking Higher Education Institutions (HEIs), has changed settings of higher education and is likely to cause further development, nationally and internationally. The rankings have gained substantial acceptance by major stakeholders –academicians, prospective students, and employers. Currently, various ranking systems are gauging and publishing the HEI's rankings, including the Shanghai Ranking Consultancy (SRC), Quacquarelli Symonds (QS), and Times Higher Education (THE). Academic rankings are facing criticism because of various issues, including the use of data sources that are not publicly available, subjective parameters, a narrow focus on research productivity and regional biases, etc.

This research work is intended to enhance the creditability of the ranking process, by using quantitative indicators based on publicly verifiable data sources. The proposed ranking methodology — OpenRank, drives the objective indicators from two well-known publicly verifiable data repositories — the ArnetMiner and DBpedia. The resultant academic ranking reflects common tendencies of the international academic rankings published by THE, QS, and SRC. Evaluation of the proposed methodology advocates its effectiveness, transparency, and quick reproducibility.

3.1. Heading Towards Objective and Publicly Verifiable Rankings

Educational institutions, especially those dealing with higher education, are considered catalysts of the national and global development process. As, they enhance a nation's competitive advantage and stimulate the development process by nurturing research, innovation, and commercialization. While ranking the HEIs, usually the ranking systems intend to measure the research productivity and teaching quality.

Some other rationales for the ranking process are summarized here. a) The rankings help HEIs in leveraging their competitive advantage in the education sector, by comparing their academic quality and research productivity with those of similar institutions at local, regional, and global levels. b) HEIs' senior management employs the ranking score, in academic policy and the decision-making process. c) Better ranking score helps an HEI in winning performance-based funding, accreditation reviews, academic program reviews, and national benchmarking processes. d) Subject-specific rankings substantially support prospective students in selecting university-major subjects [5, 6].

The process of ranking HEIs involves gathering, weighing, and analysis of the academic performance indicators. These indicators are selected and weighted according to the objectives pre-defined by the ranking entity. The ranking process has gained substantial acceptance among the major stakeholders – academicians, prospective students, and employers. They prefer an HEI with a high score in the well-known rankings [49]. Academic rankings have gained overwhelming popularity in academia and their influence is yet increasing. Nevertheless, academic experts have shown considerable reservations about the validity and transparency of the rankings. These reservations are described in section 1.3. A sample of HEIs' rankings in the year 2017, having significant differences, by the three reputed international academic ranking systems (for detail, consult section 3.2) is presented in Table 3.1. Major causes of such discrepancies are the use of data sources that are not accessible publicly and employing subjective rankings indicators. Keeping in mind the above-mentioned issues, the research problem can be posed as follows:

"Major stakeholders and end-users of the academic rankings need the academic rankings, based on objective indicators and publicly-verifiable data sources, with broader coverage of the academic institutions."

Finding a suitable solution for all the issues is currently not feasible. Nevertheless, one can suggest a more satisfying solution, for the improvement of the ranking process. This research work is intended to make the following contributions:

- Enhancing the creditability of the ranking results by merely employing the quantitative performance indicators extracted from public data sources.
- Developing an academic ranking methodology based on quantitative indicators using two well-known data repositories, DBpedia and ArnetMiner.
- Demonstrating the effectiveness of the proposed ranking methodology on real data sources.

Keeping in view the contributions, the philosophy behind the OpenRank academic ranking methodology can be summarized in the following three points.

Table 3.1 Differences in the rankings of some reputed ranking systems

Higher Education Institution	ARWU	THE	QS
Australian National University	97	48	22
Carnegie Mellon University	80	23	58
City University of Hong Kong	201	201	55
Hong Kong University of Science and Technology	201	57	36
Korea Advanced Institute of Science and Technology	201	89	46
Peking University	71	29	39
Tsinghua University	48	35	24
University of California At Berkeley	5	10	28
University of New South Wales	101	114	49
University of Sydney, Australia	83	60	46

First, the methodology emphasizes the use of quantitative ranking parameters, as they are more reliable and scientifically verifiable. These two characteristics make them less controversial. Second, the ranking methodology emphasizes the use of ranking data sources that are publicly accessible; such data sources help in producing more transparent HEIs' rankings, and their user can drill down the holistic ranking score up to the desired level. Third, the methodology emphasizes employing the ranking indicators that apply to a substantial number of global HEIs to maximize their coverage.

The rest of this chapter is organized in this way. Section 3.3 formulates the research problem, Section 3.4 presents the proposed methodology, and Section 3.5, represents the ranking results and evaluation. The research work is concluded in Section 3.6.

3.1. The Proposed Methodology

Rankings are not a new phenomenon, although their recent proliferation is, especially in the domain of HEIs. Multiple ranking systems are gauging and publishing the academic performance of HEIs worldwide. A detailed description of the well-known academic ranking systems is given in Section 1.7. Their methodologies revealed their inclination toward the potential stakeholders.

This research work is intended to enhance the creditability of the process by ranking the global HEIs, using publicly verifiable data sources and quantitative indicators. Formally, the research problem is articulated at the end of Section 2.1. Employing quantifiable ranking indicators and usage of publicly verifiable data sources, are the primary concerns of the OpenRank methodology. Nevertheless, it gives due attention to the experts' opinions and common tendencies of the well-known academic ranking systems. It drives the significant ranking dimensions and set their weights according to their worth as per the experts' judgment. However, while selecting indicators (using the Principal Component Analysis), it observes their limitations [50]. A detailed description of the methodology is following.

3.1.1. Data Sources

A major criticism faced by existing academic ranking methodologies is the use of data sources that are not publicly verifiable [7, 14]. As there is no purpose-built public data repository that would serve the intended academic ranking methodology. Consequently, it was decided to employ ArnetMiner and DBpedia repositories, as they satisfy the desired conditions. A detailed description of these two data sources is provided in Section 1.8.

3.1.2. The OPENRANK Methodology

The OpenRank methodology extracts the values of ranking indicators using the selected datasets against the following four categories — Research Productivity (RP), Academic Quality (AQ), Academic Sustainability (AS), and International Factor (IF). These ranking categories are well-known in the academic sector and used by reputed global academic ranking systems i.e., ARWU, QS, and THE university rankings [4, 6, 7]. Since the OpenRank employs multiple datasets, so the ranking categories may take their parameter values from different datasets. In our case, the Research Productivity (RP) and Academic Quality (AQ) take parameter values from the ArnetMiner dataset, while the other categories Academic Sustainability (AS) and International Factor (IF) use the DBpedia dataset. For the formal definition and explanation of these indicators, please have a look at Table 3.4.

It is also worth mentioning that the OpenRank methodology adopted a new name for the academic ranking category, "*Academic Sustainability*" because it employs some non-conventional parameters available in the datasets. For example, the Academic Age, Academic Diversity, and Financial Strength. These parameters convey meaningful information about an HEI. Nevertheless, the constituent parameters of the ranking category are being used by THE ranking system as well. The Times Higher Education (THE) employs ranking parameters "Research income" with a weight of 5.25% and "Income per academic" with a weight of 2.25%. Please consult Table 1.5 (THE Ranking Methodology). he weights (w_1, w_2, w_3, and w_4) of indicators in the OpenRank methodology are assigned keeping in view the tendencies of the benchmark academic ranking systems – ARWU, THE, and QS academic rankings.

A comparison of various academic ranking dimensions used by the benchmark ranking systems is presented in Table 3.2. The weights used for the ranking dimensions show their importance for the ranking entities. According to the comparison, the ranking methodology of THE seems more research-oriented, whereas the ranking methodology of the QS shows a balanced inclination toward research productivity and academic quality. Nevertheless, it also gives substantial weight (40 percent) to a qualitative ranking parameter – the *Academic Peer Review*. The ranking methodology of ARWU is focused on *Research Productivity* and *Academic Quality*. Nevertheless, its inclination is toward *Academic Quality* (with 50 percent weightage).

For instance, the ranking dimension *Research Productivity* has 65, 20, and 40 percent in methodologies of THE, QS, and ARWU whereas in OpenRank methodology it has 45 percent weight – approximately average of the weights used in other methodologies. The same is the case in the weight of *Academic Quality*. Some of the parameters (i.e., *Research Student Ratio, Favorite Venue Publications,* and *Academic Diversity*) employed by the OpenRank methodology are not exactly common in other methodologies. Nevertheless, they are highly relevant to the analysis and communicate important information about an HEI. These parameters are directly or indirectly related to the core ranking dimensions commonly employed in the three methodologies.

Overall, the OpenRank methodology employed fifteen ranking indicators in the four categories. Table 3.3 lists ranking indicators of the OpenRank along with their weights and descriptions (acronyms described). Similarly, Figure 3.1 depicts the workflow of the OpenRank methodology.

The ranking metric is modeled as follows:

Rank (HEI) = w_1AQ + w_2RP + w_3AS + w_4IF

where $w_1 = 0.35$, $w_2 = 0.45$, $w_3 = 0.15$ *and* $w_4 = 0.5$

AQ = 8TAS + 7TS + 10RFC + 10FSR

RP = 5RFR + 5HCC + 5RSR + 5CPF + 10HPR + 15FVP

AS = 3AA + 7AD + 5FS

IF = 3IFR + 2IS

Here the ranking metric is formulated mathematically:

$$R_D = \{ RP, AQ, AS, IF \} \tag{1}$$

Let R_D be the ranking dimensions i.e. Where the elements (RP, AQ, AS, IF) represent Research Productivity, Academic Quality, Academic-Sustainability, and International-Factor, respectively. Where the elements (RP, AQ, AS, IF) represent Research Productivity, Academic Quality, Academic-Sustainability, and International-Factor, respectively. Each element (indicator) in the set RP_i is an ordered pair of an indicator's value and its corresponding weight '*w*'. Therefore, it is equated as:

$$RP_{ij} = \{ (rp_i, w_j) \} \tag{2}$$

where rp_i is the value of the indicator and w_j is its corresponding weight. Likewise, equations for the rest of the elements of R_D can be represented as:

$$AQ_{ij} = \{ (aq_i, w_j) \} \tag{3}$$

$$AS_{ij} = \{ (as_i, w_j) \} \tag{4}$$

$$IF_{ij} = \{ (if_i, w_j) \} \tag{5}$$

The aggregate of the indicator's weighted values of a ranking category can be computed using the following equation.

$$s(rp) = \sum_{rp \in RP_i}^{N_{rp}} (rp * w) \tag{6}$$

where N_{rp} is the total number of indicators in the ranking category (dimension) RP Likewise, equations for the rest of the elements in the set R_D can be derived. The aggregate Score (Rank) of an HEI can be obtained by adding scores obtained by it, against the individual dimensions using the following equation.

$$Rank(HEI) = \sum_{d=d_i}^{k} s(d_{HEI}) \tag{7}$$

where k = | R_D | and d_{HEI} is the ranking category (dimension) for the respectiveHEI. For the calculation of Equation 7, the following algorithm was used.

3.1.3. Performance Evaluation

To compare the results of the OpenRank methodology, the two well-known and widely accepted academic ranking systems — ARWU and QS were used as the baseline measures. Ranking methodologies of the baseline measures are given in Table 1.4 and Table 1.5. Their ranking results for the year 2017 are used for comparison.

The Overlap (O) and Average Overlap (AO) similarity measures [51, 52], are commonly used for non-conjoint ranking lists. The measures gave an intersection of the ranking lists of the OpenRank and baseline academic ranking systems.

Mathematically the evaluation measures are expressed as:

Overlap
$$O(L_1, L_2)_N = \frac{|\ L_{1(N)} \cap L_{2(N)}\ |}{N} \tag{8}$$

Average Overlap
$$AO(L_1, L_2)_N = \frac{1}{N} \sum_{i=1}^{N} \frac{|\ L_{1(N)} \cap L_{2(N)}\ |}{N} \tag{9}$$

where $O(L_1, L_2)_N$ and $AO(L_1, L_2)_N$ are the Overlap and Average Overlap similarity measures, respectively.

3.2. Experimental Setup

To implement the proposed ranking methodology, the two datasets were processed differently. For scoring the ranking indicators based on the ArnetMiner dataset, was downloaded, and processed its latest version (ACM-Citation-Network V8, released in April 2016) on the local machine. The dataset was created for social network analysis and identification of connections between researchers, conferences, and publications. Nevertheless, after some careful processing, it provided valuable HEIs' performance indicators. Some of the derived academic ranking indicators are *Highly Productive Researchers, Influential Venue Publications, HEI's Citation Count, Citation Per Faculty, and Researcher-Faculty Ratio,* etc. as listed in the data was processed through its SPARQL Endpoint using the SPARQL queries, to score ranking indicators based on the DBpedia dataset. DBpedia provided ranking indicators like *Total Academic Staff, Total Students, Academic Age, Financial Strength, International Factor,* etc. For brevity, Table 3.5 shows the ranking scores of top-10 HEIs, against the categories. The full list of the top 100 HEIs ranked according to the methodology, is provided as a part of the appendix. To determine the similarity between the OpenRank ranking with those produced by the baseline ranking systems, the Overlap Similarity and Average Overlap Similarity measures were employed. For better insight, the Average Overlap Similarity was calculated for the top fifty HEIs in five intervals, each containing ten HEIs.

3.3. Results and Discussion

It is worth mentioning that although, the academic data available on the Internet is abundant and perpetually increasing. Nevertheless, most of the data repositories are not purposely built for academic ranking. The same is the case with DBpedia and Arnet Miner repositories. These data sources were employed because of their high relevance to the intended analysis and the ease of access they offer. So, data quality is an issue that is hard to tackle. In the future, the Open-Rank methodology would produce more reliable and optimal results by employing more enriched and curated data sources. After scoring and weighting the indicators, the HEIs were ranked based on their aggregate score which is calculated using equation 7 through the algorithm.

Table 3.2. A comparison of the ranking dimensions in various methodologies

Ranking Dimension	**Weight in percentage (%)**			
	THE	**QS**	**ARWU**	**OpenRank**
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**

The OpenRank Algorithm

BEGIN

INPUT: IndWeight, HEI[RankInd]

OUTPUT: Rank [HEI]

BEGIN

1. **for each** hei ∈ HEI
2. Rank[hei] = 0
3. **for each** ri ∈ RankInd
4. Rank[hei] += [hei][ri] * IndWeight[ri]
5. **end for**
6. **end for**
7. return Rank [HEI]

END

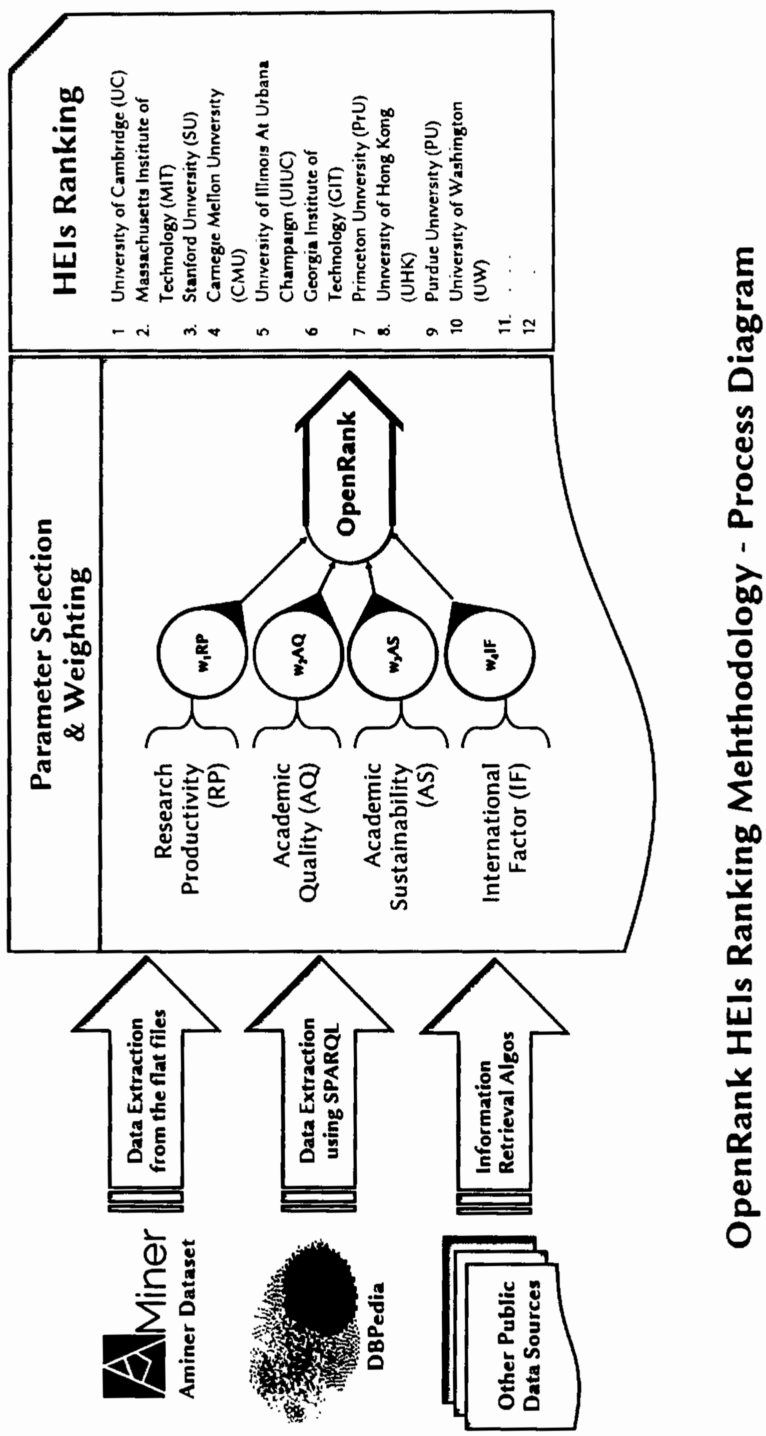


Figure 3.1. Workflow of the OpenRank Methodology

Table 3.3. Ranking Indicators of the Proposed Methodology

	Indicator	Weight (%)	Description
	Research Productivity (RP)	**45**	
1	Researcher-Faculty Ratio (RFR = RFC/TAS)	**5**	The ratio of researchers (RFC) to the total academic staff (TAS).
2	HEI Citation Count (HCC)	**5**	Total no. of citations earned by the researchers affiliated with an HEI.
3	Research Student Ratio (RSR)=PGS / (PGS+UGS)	**5**	The ratio of regular post-graduate students to undergraduate students.
4	Citation Per Faculty (CPF)	**5**	Total citations (HCC) to the total academic staff (TAS).
5	Highly Productive Researchers (HPR)	**10**	HEI's Influential researchers have more than 250 publications, in the ArnetMiner dataset.
6	Favorite Venue Publications (FVP)	**15**	No. of publications by an HEI, on the Journals & conferences having more than 3500 publications.
	Academic Quality (AQ)	**35**	
7	Total Academic Staff (TAS)	**8**	Total number of permeant faculty members
8	Total Students (TS)	**7**	Total number of currently enrolled students
9	Research Faculty Count (RFC)	**10**	Count of the research faculty affiliated with an HEI and listed in the ArnetMiner citation database.
10	Faculty-Student Ratio (FSR=TS/TAS)	**10**	A measurement of teaching commitment
	Academic Sustainability (AS)	**15**	
11	Academic Age (AA)	**3**	Academic age of an HEI in years
12	Academic Diversity (AD)	**7**	Count of the internationally well-known academic faculties owned by an HEI.
13	Financial Strength (FS)	**5**	Endowment Fund \ Annual Budget of an HEI, in US Dollars.
	International Factor (IF)	**5**	
14	International Faculty Ratio (IFR)	**3**	The ratio of the foreign qualified faculty and local faculty
15	International Students (IS)	**2**	The ratio of the full-time foreign students to the local students in an HEI.

Table 3.4. Overlap & Average Overlap similarity (% age) with the baseline ranking measures

Ranking Systems	Overlap Similarity	Average Overlap Similarity
ARWU	79	74
QS	85	82

The similarity gradually decreases in the first five intervals. A similar trend can be seen between the results of the QS and OpenRank results, although in this case, the number of matching HEIs is fewer. The higher similarity between the results of the ARWU and OpenRank than those of QS is satisfactory as the academic experts consider ARWU more quantitative than the QS rankings. The overlap similarity would be further improved by using more open and curated data sources. The currently employed data sources have been created by automatic or semi-automatic extraction of content from semi-structured or unstructured sources, therefore lack of completeness and consistency is highly likely.

The use of the ArnetMiner dataset, enabled the OpenRank methodology to explore new ranking indicators like the count of publications of an HEI on the favorite venues (top 50 venues recorded in the dataset, having more than 3500 publications). Similarly, *HEI's Citation Count* and *Highly Productive Researchers* reveal new insight into the ranking process. Figure 3.4 represents the count of the Highly Productive Researchers (HPR) and publications on the Influential venues (FVP) for the Top-10 HEIs.

As the employed datasets do not manage the data with the year-time stamp; so, it was necessary to use the cumulative sum. Yes, these thresholds are achieved by optimizing coverage of the reasonable number of "Highly Productive Researchers" (HPR) and the "Favorite Venues for Publications" (FVP). Taking a higher number was causing the exclusion of many HPRs and FVPs. Due to the incompleteness of the data sources, some institutions may be under-represented in this category. Although measuring the count of the publications by an HEI on well-known publication venues is important, nevertheless, the collection of complete data is a limitation faced by the methodology.

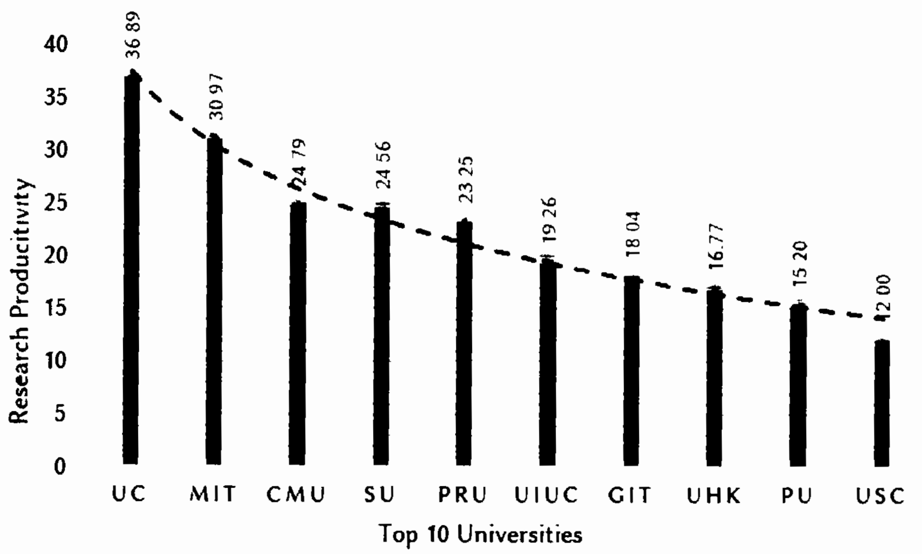


Figure 3.2. Top 10 HEIs w.r.t Research Productivity

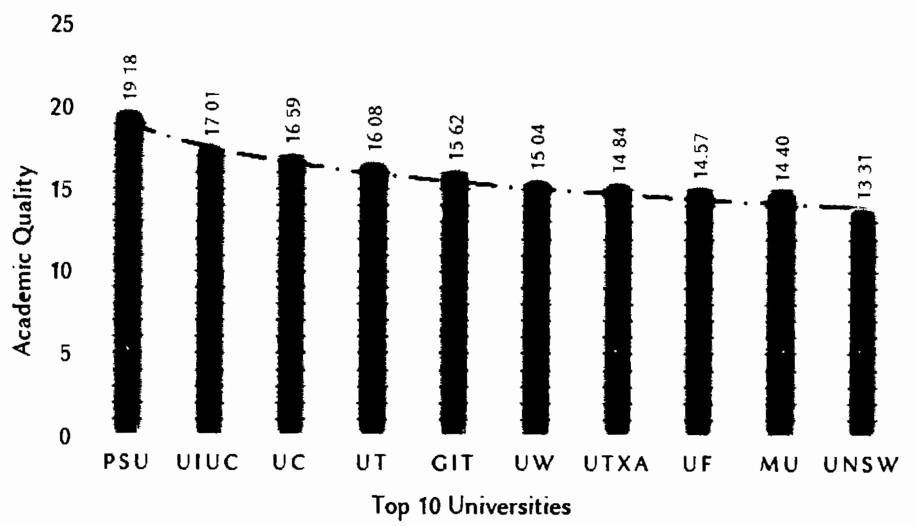


Figure 3.3. Top 10 HEIs w.r.t Academic Quality

For the full name of an HEI against its abbreviation, please consult the ranking results provided in the appendix.

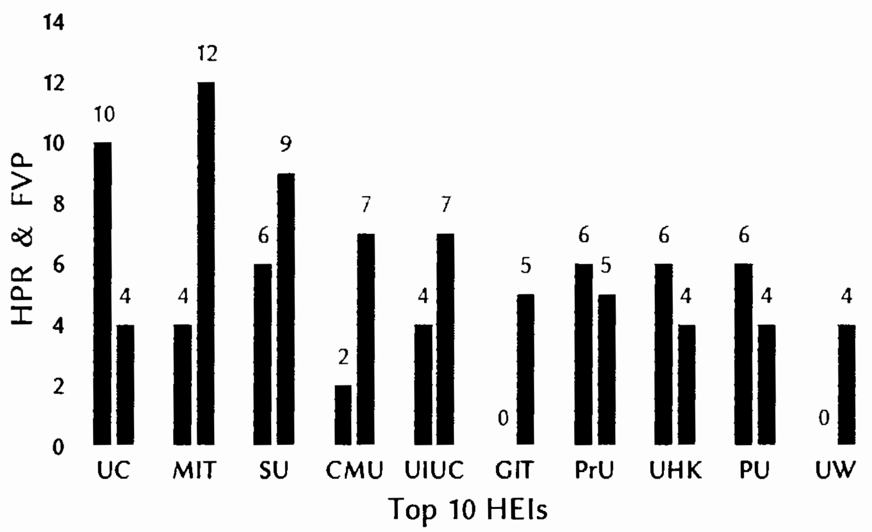


Figure 3.4. Highly Productive Researchers (HPR) and Publications on Favorite Venues (FVP)

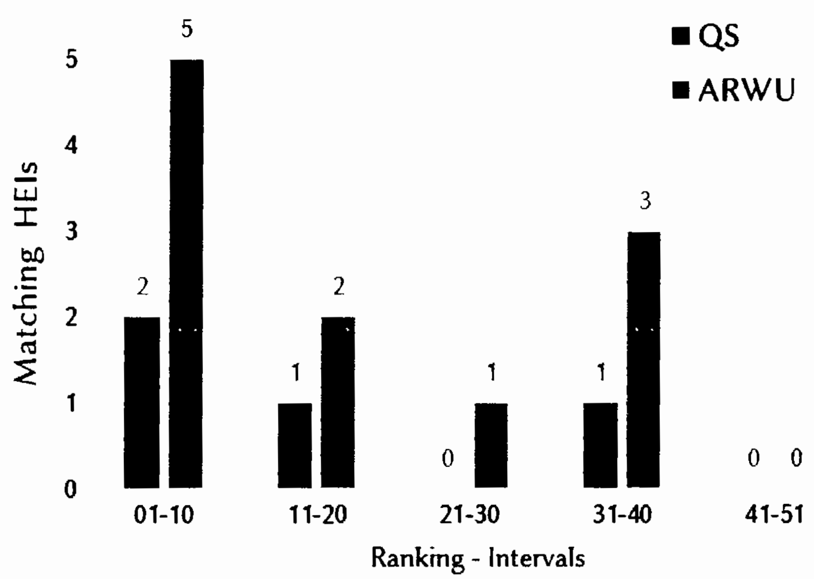


Figure 3.5. Count of Matching HEIs (Top 50) among the OpenRank, ARWU, and QS

The DBpedia data repository contains information about the geographical regions and countries of the world. The methodology gauged the HEI's share in the regions (Asia, Europe, Oceania, and North America) and countries as well. Figure 3.9 presents the region-wise count of the top 100 HEIs while Figure 3.8 highlights the countries having a significant number of HEIs having excellent performance.

The OpenRank methodology explored the relationship between the Human Development Index (HDI) of a country and the number of HEIs owned by it. Figure 3.6 plots the relationship for the countries having a significant number of HEIs. Interestingly, the trend highlights the fact that outstanding human prosperity does not ensure an outstanding share in research and innovation – nevertheless, the United States of America and the United Kingdom are proving the need for human prosperity for better performance. It is also interesting to represent the relationship between the GDP of a country and the number of top-performing HEIs owned by it.

The trend reveals a direct proportion with the Grand Domestic Product (GDP) of a country and the number of HEIs owned by it. The United States of America and the United Kingdom are once again outperformed in having top-performing HEIs, as compared to various other countries with a higher GDP. Overall, the proposed ranking methodology reflected the research objectives — ranking HEIs using objective parameters and publicly verifiable data sources. Quick and cost-effective data acquisition is an additional benefit of the methodology. The resultant ranking is evaluated and equated with rankings produced by the two reputed academic ranking systems, ARWU rankings, and QS University rankings in 2017. The ranking scores of top-10 HEIs, against the four ranking categories, are shown in Table 3.5.

Interestingly, the HEI's rankings are different w.r.t the Overall Score, Research Productivity, and Academic Quality as shown in Table 5, Figure 3, and Figure 4. This difference demands uncovering the fine-grained research and academic excellence of HEIs, as suggested in [28].

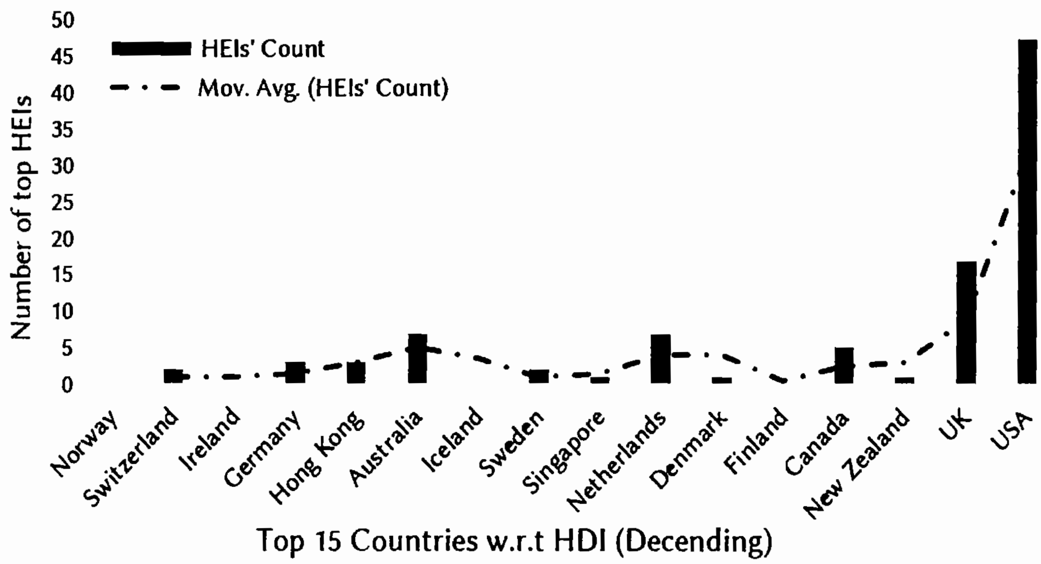


Figure 3.6. No. of HEIs Versus Human Development Index (Descending)

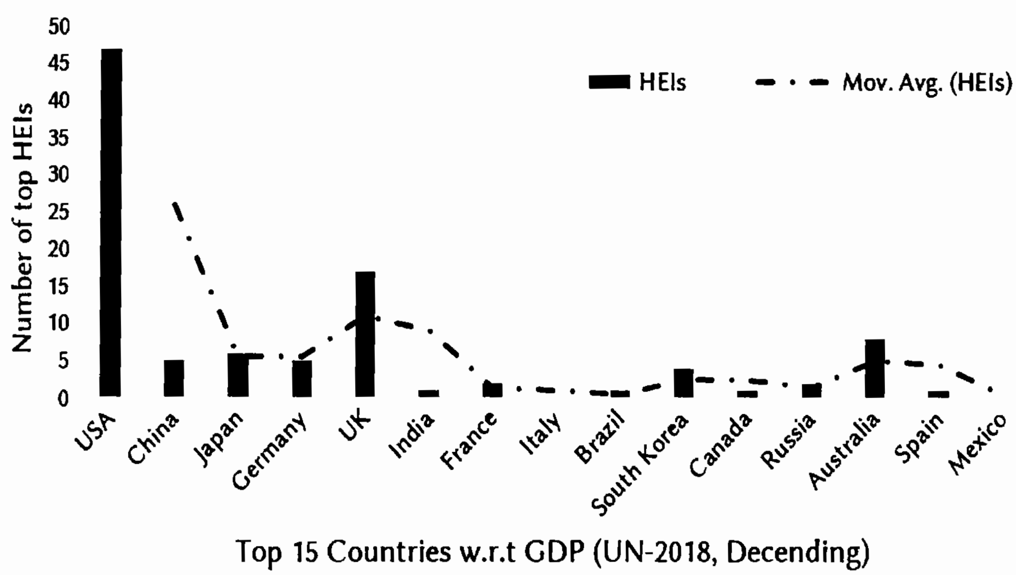


Figure 3.7 No. of HEIs Versus Country Grand Domestic Product (Descending)

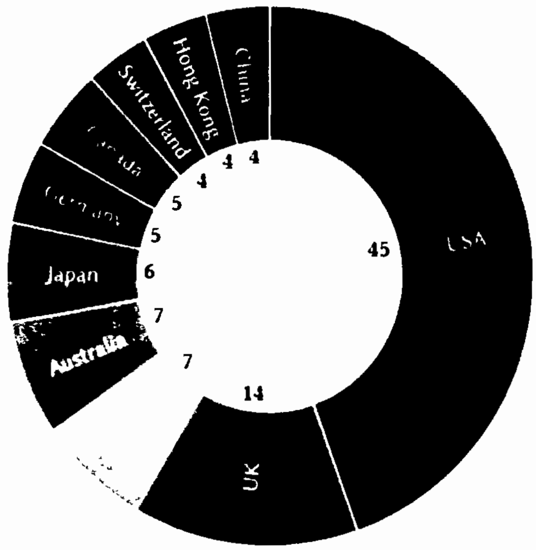


Figure 3.8. Countries having a significant number of top HEIs

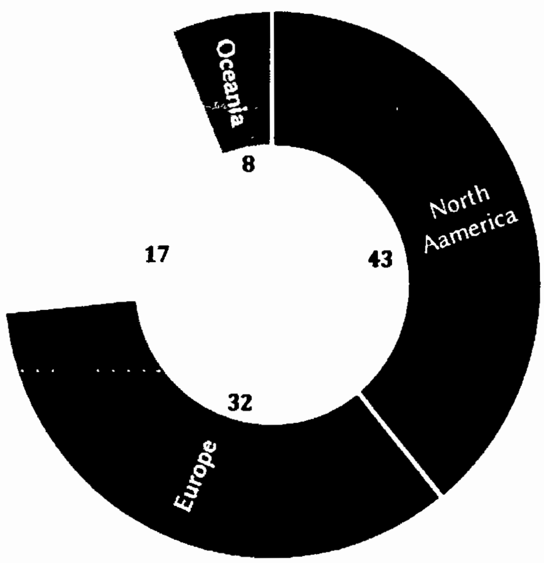


Figure 3.9. Regions having a significant number of top HEIs

Table 3.5. Top -10 Higher Education Institutions, as ranked by the Open Ranking

	Higher Education Institution	Research Productivity	Academic Quality	Academic Sustainability	International Factor	Overall Score
1	University of Cambridge (UC)	36.9	16.6	10.1	2.7	66.27
2	Massachusetts Institute of Technology (MIT)	31.0	13.9	10.3	3.2	58.40
3	Stanford University (SU)	24.6	11.2	10.8	2.6	49.06
4	Carnegie Mellon University (CMU)	24.8	10.4	9.1	2.7	47.03
5	The University of Illinois At Urbana Champaign (UIUC)	19.3	17.0	9.2	1.0	46.49
6	Georgia Institute of Technology (GIT)	18.0	15.6	9.6	0.8	44.07
7	Princeton University (PrU)	23.3	7.2	9.8	1.6	41.87
8	University of Hong Kong (UHK)	16.8	10.8	9.5	3.6	40.71
9	Purdue University (PU)	15.2	12.9	8.6	0.2	37.00
10	University of Washington (UW)	10.0	15.0	10.2	1.4	36.65

The results of Overlap Similarity and Average Overlap Similarity, among the OpenRank and the baseline measures, are presented in Table 3.4. For top-50 HEIs, the Average Overlap Similarity (AOS) among the ranking lists of OpenRank, ARWU, and QS is calculated in five intervals in Figure 3.5. The similarity analysis reveals the closeness of the ranking results produced by OpenRank and ARWU.

3.4. Summary

Academic rankings are facing criticism because of various issues, including the use of data sources that are not publicly available, subjective parameters, a narrow focus on research productivity and regional biases, etc. In this chapter, the differences among the results of the academic rankings published by ARWU, THE, and QS are analyzed.

To address some of the causes of the differences, a new academic ranking methodology – OpenRank was developed. It was based on the quantitative indicators extracted from publicly verifiable data sources. The methodology demonstrated its potential to enhance the credibility of the academic ranking process, by successfully ranking the global HEIs, using two publicly verifiable data sources — *DBpedia* and the *ArnetMiner*. In the methodology, both research productivity and academic contribution of an HEI get reasonable weights. This comprehensive approach broadens coverage of HEIs, especially for newly established universities. Evaluation of the methodology advocates its effectiveness, transparency, and quick reproducibility.

Results of the OpenRank methodology are also reported in a research article [35], that has been published in a well-known research journal. A research recommender system for novice researchers and prospective students is recommended by employing such publicly available data sources. The need for a purposely built, publicly verifiable electronic data source, for performance evaluation of the global HEIs, is also among the future implications of this research effort.

Chapter 4

Heading Towards the Fine-Grained Academic Rankings

The academic ranking process has considerably evolved in the past fifteen years. 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 researchers' profiles from the Google Scholar Citations repository. While the DBpedia repository is employed for 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 concerning 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 ARUW, THE, and QS rankings systems. The growing interest in subject-specific and sub-discipline-specific rankings is irreversible. The fine-grained analysis is a response to the need.

4.1. Introduction

The role of Higher Education Institutions (HEIs) as a catalyst of the national and global development process, has become established worldwide. Governments of developed countries nurture research, innovation, and commercialization through 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.

The first academic ranking of global HEIs was published in 2003; today the ranking outcome has gained considerable acceptance by the major stakeholders [50, 35]. Bibliographic data repositories are reporting a significant increase, in the publication of research articles focused on the academic ranking process. According to a study [2], 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. The increasing interest in academic rankings is understandable because of its various applications as discussed in [35, 53]. Nevertheless, the ranking results need to be more transparent and fine-grained, as highlighted in [37, 54, 35]. Major issues of the ranking process are summarized in Section 1.7.

During the intended fine-grained analysis, the holistic ranking scores allocated to the top-performing HEIs in the years 2017, 2018, and, 2019 by the three reputed ranking systems – ARWU, THE, and QS were examined. To obtain a more stable ranking score for each of the HEI, the 3-year average of the ranking scores, for each of the systems was computed. For descriptive analysis, the average of the nine ranking lists (Grand Average Rank – GAR Score) was also computed. It was produced by the three ranking systems in the stipulated years. A summary of the analysis is presented in Table 4.1 and a part of the results is visualized in Figure 4.3 and Figure 4.4.

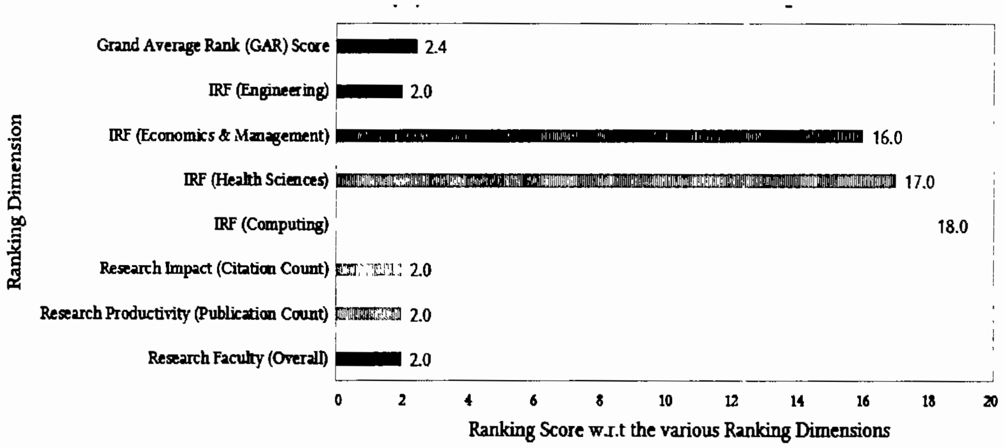


Figure 4.1 Stanford University positions w.r.t various ranking dimensions

For instance, the USA-based Stanford University (SU) stands first according to the GAR score. In other words – unanimously ranked first position, by ARWU, THE, and QS in the years 2017, 2018, and 2019. Nevertheless, the university attains significantly different ranks according to (w.r.t) the various ranking dimensions. For example, as shown in Figure 4.1, 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 Sciences, and Computing, respectively. Nevertheless, the rank of the university w.r.t the research productivity, 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 4.1. This analysis also gave multiple insights into the subject-specific rankings of HEIs in the four scientific disciplines.

4.1.1. Research Objectives

The use of 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 [21] 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. A research effort aimed at enhancing the credibility of the ranking process by ranking the global HEIs using objective ranking indicators that were extracted from publicly verifiable data sources. This work was reported in [35]. The *OpenRank* methodology suggested in the research was developed for the holistic ranking of the HEIs. Nevertheless, during the research effort, the need for a fine-grained analysis at the subject-specific level was seriously felt.

4.1.2. 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 and Google Scholar's Citations.
- 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 Section 4.2 presents the proposed ranking methodology. The rankings measures are presented in Section 4.3. The experiments and their results are reported in Section 4.4 and Section 4.5 Finally, Section 4.8 concludes the research work.

4.2. The Ranking Methodology

Major issues of the ranking process as highlighted in the literature [37, 54, 35], are summarized in Section 1.7. This research work focuses on the fine-grained analysis of academic data. It 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 analysis using publicly verifiable data. 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-100s in the three stipulated years, by the three well-known rankings systems including the ARWU, THE, and the QS.

Table 4.1. HEIs' Average Rank in 3 years by the three ranking systems and GAR Score

	Higher Education Institutions (HEIs)		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	Up	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 Zurich	ETHZ	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

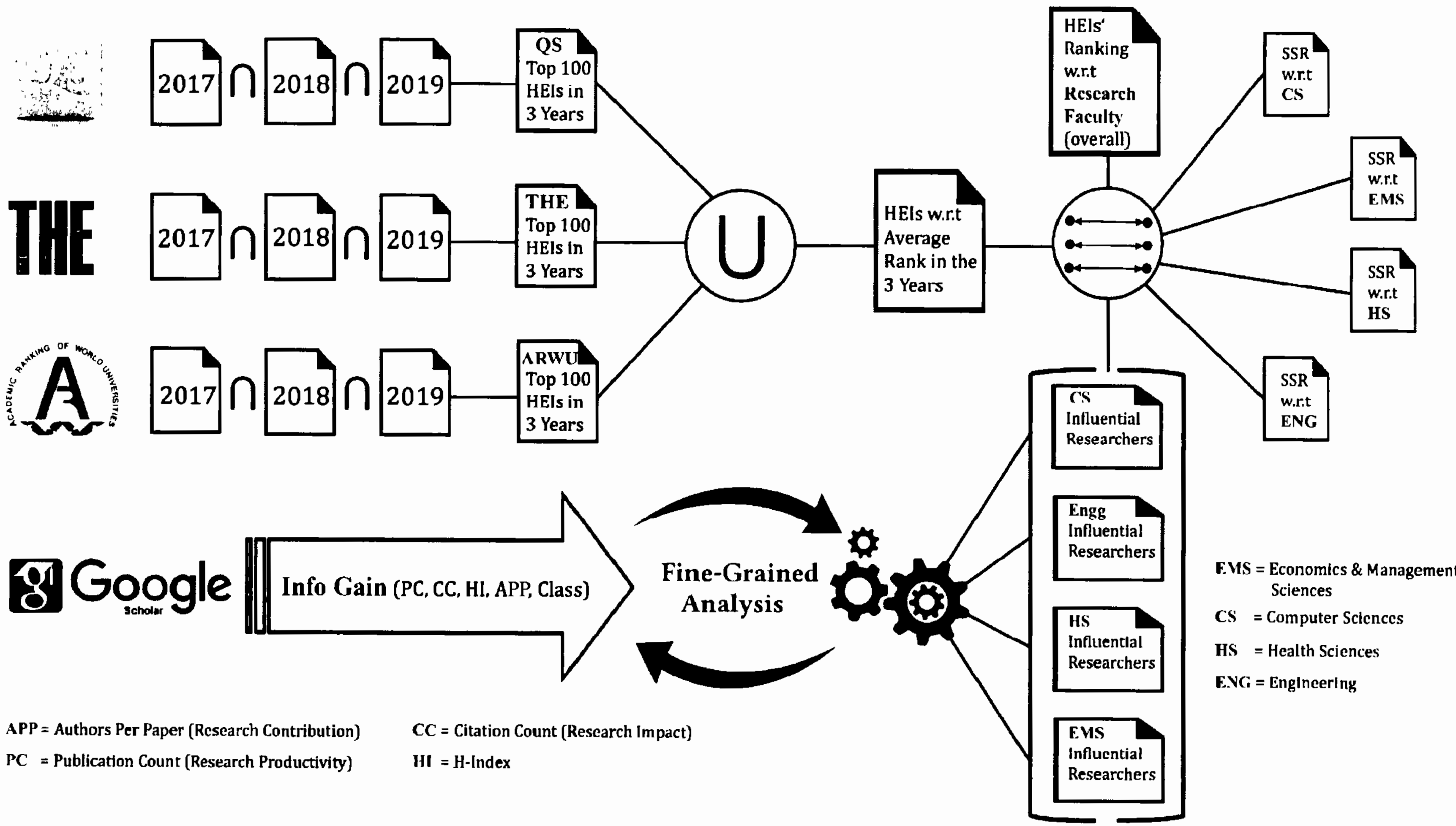


Figure 4.2 Workflow of the OpenRank Methodology

4.3. Data Sources

A major criticism faced by existing academic ranking methodologies is the use of data sources that are not publicly verifiable [37, 14]. Keeping in view the research objectives, it was decided to employ multiple data sources that satisfy the necessary conditions. For fine-grained analysis of the academic ranking statistics, the existing ranking results were reused. They were produced by the three well-known ranking systems — ARWU[6], THE[7], and QS[8]. These results are publicly available on their official websites. Some of the systems provide a limited amount of data in spreadsheets or flat files. Various statistics about influential researchers were extracted from Google Scholar Citation (GSC) [55]. 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. By default, GSC orders researchers' profiles according to their citation count. A brief description of the information extracted from GSC is presented in Table 4.3. For more detail about the data repository, please consult Section 1.8.

Adopting a careful data reduction approach, only the top two hundred researchers' profiles were selected w.r.t the *Citation Count* (or research impact). To ensure a substantial *Publication Count* (or research productivity), the profiles having at least 100 publications were filtered. For a more generalized analysis, the profiles from the four scientific disciplines – Computer Sciences (CS), Health Sciences (HS), Engineering Technology (ET), and Economics and Management Sciences (EMS) were selected. 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, other scientific disciplines would be considered as well. The intended data analysis is based on the GSC data snapshot taken in November 2019.

---

[6] http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/index html [Accessed on August 2021]

[7] https.//www timeshighereducation com/world-university-rankings/2021/world-ranking [Accessed on August 2021]

[8] https //www.topuniversities com/subject-rankings/2020 [Accessed on November 2020]

The DBpedia repository is maintained by Wikipedia. It is among the significant data repositories in the LOD cloud [56, 57]. The cloud contains structured data related to various domains like Geography, Healthcare, Media, and Academia [57]. Meaningful information about the HEIs, researchers and research publications can be obtained from the cloud. In this research work, the DBpedia repository was employed by posing the SPARQL [56, 58] 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[9].

4.4. The Ranking Measures

Heading toward the fine-grained analysis of the academic performance evaluation data, a systematic approach was adopted. 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 reputed 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, the HEIs were determined that repeatedly maintained their position in the list of top-100, for 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, the union of the three lists was calculated. This gave the final list of top-performing HEIs in the stipulated three years. To obtain a stable ranking score for an HEI, the 3-year average of the raking score was computed, for each of the three ranking systems. A comparison of the 3-year average scores also highlighted the significant differences among the ranking systems even for a single HEI. The Grand Average Rank (GAR score) was also calculated, to get 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.

[9] https://github.com/muhammadsajidqureshi/OpenRank [Accessed on Jul 2019]

Table 4.2. A comparison of the ranking dimensions in various methodologies

Ranking Dimension	Weight in percentage (%)			
	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**

Table 4.3. The researchers' 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

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, the top 20 HEIs were focused, 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 Table 4.1.

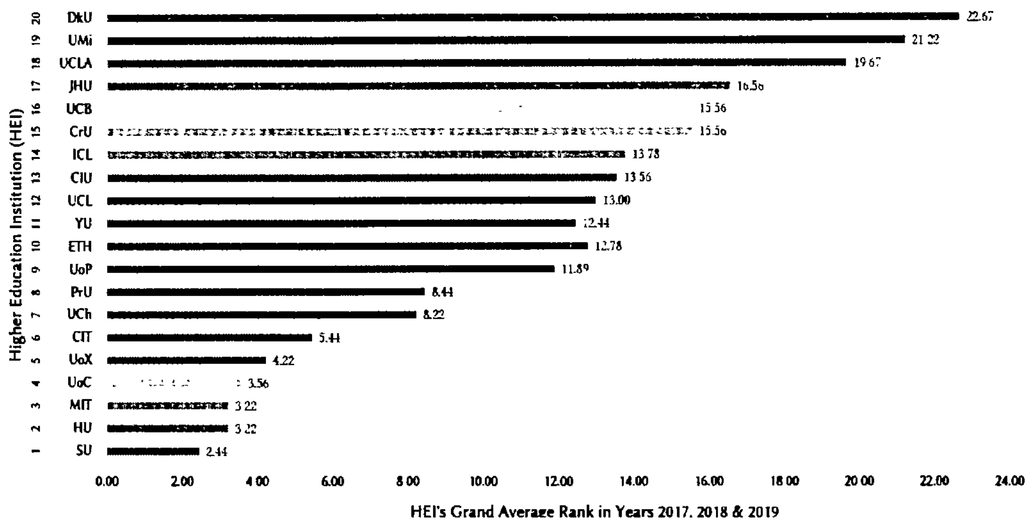


Figure 4.3. Top 20 HEIs w.r.t the GAR Score, in Years 2017—2019

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 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 Equation 1.

$$\text{Top100HEIs} = \bigcup \left(\forall \text{ RankEnt} \bigcap_{Y=2017}^{2019} \text{Top100HEIs}_{\text{RankEnt}} \right) \qquad (1)$$

where RankEnt $\in$ { ARWU, THE, and QS }

The profiles of influential researchers for the four scientific disciplines were extracted from the GSC repository. For fast data processing, only two hundred researchers were selected for each of the disciplines, using a threshold of hundred publications (PC => 100). The mathematical equations for the four sets { Res_{cs}, Res_{hs}, Res_{ems}, Res_{eng} } containing the researchers' profiles are given below.

$$R_{cs} = \{ R_{cs} \mid R_{cs}(PC) \geq 100 \} \qquad (2)$$
$$R_{hs} = \{ R_{hs} \mid R_{hs}(PC) \geq 100 \} \qquad (3)$$
$$R_{ems} = \{ R_{ems} \mid R_{ems}(PC) \geq 100 \} \qquad (4)$$
$$R_{eng} = \{ R_{eng} \mid R_{eng}(PC) \geq 100 \} \qquad (5)$$
$$R = R_{cs} \cup R_{hs} \cup R_{ems} \cup R_{eng} \qquad (6)$$

Overall, the profiles of eight hundred researchers were retrieved by taking unions of the four sets, as expressed in Equation 6. GSC maintains various valuable statistics about the researchers and millions of their research publications. In the research analysis, various statistics were extracted from the repository, 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 – APP). The title, publication venue, and names of the co-author(s) were also retrieved for the research publications published by the researchers. A brief description of these attributes is summarized in Table 4.3.

The dataset was used 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 experts' opinions, these parameters are treated as the Key Performance Indicators (KPIs) in the related literature [15, 18]. Second, most of the bibliographic data repositories essentially record them. An overview of the meta-data of GSC, DBLP, Microsoft Academics, and the ArnetMiner data repositories, confirms the fact [21]. The metrics themselves are also self-explanatory [32]. 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 [59, 60]. It is the count of the citations of a researcher's 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 and quality of work of a research scholar [61]. Although the *Hirsh-Index* faces many critics, nevertheless, the wisdom behind the measure is very foundational; therefore, the metric in its original, or its variants has been surviving for the last fifteen years, in the bibliographic data repositories [62, 63].

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 [38, 64, 65]. Nevertheless, since our research work is focused on mapping affiliations of the influential researchers with the top-performing HEIs; therefore, a simple approach is adopted by assigning fifty percent weight to the first author and dividing the remaining fifty percent, equally among the co-authors [66, 67]. 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 researchers' ranking methodology, the issue of assigning weights to the parameters surfaced. The weights used for the ranking parameters show their importance in the process of performance evaluation. It seems suitable to mention that there is no standard weighting criterion for the selected ranking indicators, therefore their weights usually remain controversial. The well-known InfoGain (IG) [32, 68] algorithm was applied to the target dataset, to determine the weights (*w1, w2, w3, and w4*) 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. The GSC default researchers' ranking criterion — the *Citation Count* (CC) was employed 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.

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

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 quantity does not ensure quality, a more comprehensive methodology was adopted to rank the research scholars. Equation 8 represents the researchers' ranking methodology with the weighted parameters.

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

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 Equation 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 researchers' profiles does not produce justifiable ranking results. Therefore, the ranking methodology was applied 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.

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

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

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

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

Overall, the dataset having ranked researchers' profiles of all the disciplines was developed according to Equation 13.

$$\text{InfRes} = \text{InfRes}_{cs} \cup \text{InfRes}_{hs} \cup \text{InfRes}_{ems} \cup \text{InfRes}_{eng} \quad (13)$$

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

$$UniRank(\ InfRes,\ TopHEIs\ ) \quad (14)$$

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 researchers' 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, it can be described as the HEIs ranking for the *Research Productivity* irrespective of the research disciplines. Nevertheless, the ranking process was drilled down and found the subject-specific ranking of HEIs w.r.t *Research Productivity* using the following set of equations.

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

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

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

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

The fine-grained analysis proceeded 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 4.6.

4.5. Experiments

The required data was acquired from the data sources introduced in Section 4.3, using multiple techniques. To obtain the list of top-100 HEIs with consistently outstanding performance in the stipulated three years, the HEI's rankings were retrieved from official web portals of ARWU, THE, and QS academic ranking systems. The profiles of influential research scholars were retrieved from the GSC data repository. To obtain the necessary data about the HEIs and countries, the DBpedia data repository was employed. The data was processed through its SPARQL Endpoint using the SPARQL queries [58]. 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, the data was verified from the websites of the HEIs. Finally, the researchers' affiliations were mapped 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 4.4. Similarly, using the data about the base countries, the countries having a considerable number of HEIs in their different cities, were also identified. Quite understandably, the top fifteen countries having HEIs with outstanding performance, are the most stable and prosperous courtiers. Figure 0.12 represents a part of the analysis. Governments of 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 [34], the argument was discussed, and the comparisons of HEIs having very different sizes, budgets, 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 a 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 the such fine-grained analysis. For a quick understanding, please consult the process diagram presented in Figure 4.2. 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 suggested in [7, 26]. The holistic ranking scores published by various ranking systems face criticism termed the "*Misinterpretation of the academic performance*" because of such discrepancies. Such discrepancies would be minimized by the subject-specific rankings or even better by ranking HEIs w.r.t various ranking dimensions such as *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.*

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 Equation 1. Then the profiles of the influential research scholars having a considerable publication count ($PC \geq 100$) were selected. They were belonging to the four disciplines. The selection criteria are modeled in the set of equations labeled Equation 2 to Equation 5. Then, the profiles were accumulated by the union operation presented in Equation 6. The researchers' ranking methodology was applied to the profiles to gauge their influential status. The mathematical expression of the methodology is presented in Equation 7.

4.6. 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. The DBpedia and GSC repositories were employed because of their high relevance to the intended analysis. In the future, the fine-grained analysis strategy would employ more relevant and curated data sources.

4.6.1. 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 Figure 4.3. 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 equal to 25 by the ARWU. ICL being among the top ten world-class institutions would be reluctant in accepting such a large difference.

4.6.2. 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, the Citation Count (CC) was extracted. It was attained by the Influential Research Faculty (IRF) affiliated with the top-performing HEIs.

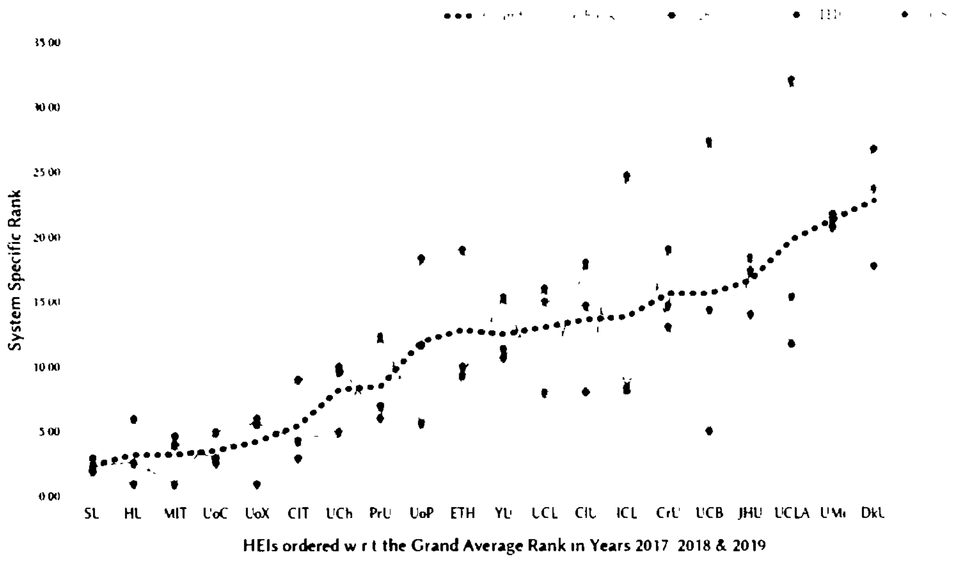


Figure 4.4. Top 20 HEIs w.r.t the 3-year Average Rankings by ARWU, THE, QS, and GAR score

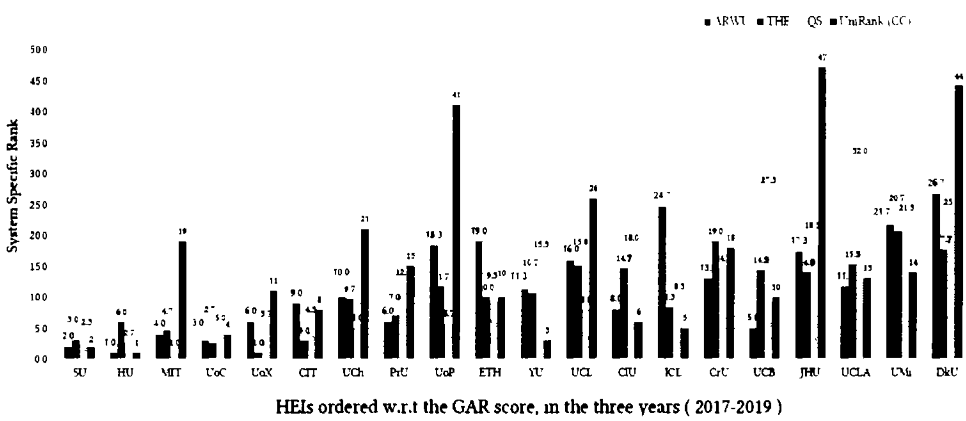


Figure 4.5. Top 20 HEIs w.r.t the GAR and the Research Impact

Figure 4.5 presents the ranking of the HEIs w.r.t the Research Impact or Citation Count (CC). 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 a fine-grained analysis of the holistic ranking scores published in the rankings lists of the reputed ranking systems.

For brevity, a comparison of top-20 HEIs is presented. 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 academic data. Heading toward a more transparent and fine-grained analysis is inevitable in the ensuing years.

4.6.3. 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. The Publication Count (PC) was also extracted and accumulated. It was recorded against the influential researchers affiliated with the top-performing HEIs. This ranking is termed the HEIs ranking w.r.t Research Productivity (UniRank). The list of the top twenty HEIs, w.r.t the research publications is presented in Table 4.6. For a better understanding, the rankings are visualized in Figure 4.6. In the figure, the green color bars represent the ranking w.r.t the Research Productivity, whereas the other colored bars represent the 3-year average ranking scores of HEIs that were allocated by ARWS, 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, the subject or discipline-specific Publication Count (PC) was also extracted. 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 statistics. In Figure 4.7, the subject-specific rank of the top 20 HEIs, is represented by the dark blue, yellow, blue, and green bars. The ranks of HEIs in the four disciplines w.r.t the research productivity were determined. The lists of the top five research scholars in the four scientific disciplines are presented in the four tables Table 4.5 Computer Science (CS), Table 4.7, Economics and Management Sciences (EMS), Table 4.9 Health Sciences (HS), and Table 4.11 Engineering (ENG). The subject-specific rankings are collectively visualized in Figure 4.7. An overview of the graph reveals the fact 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 4^{th}, 10^{th}, 8^{th}, and 1^{st} for 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 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, it can be observed that other HEIs have 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 such 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 *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.*

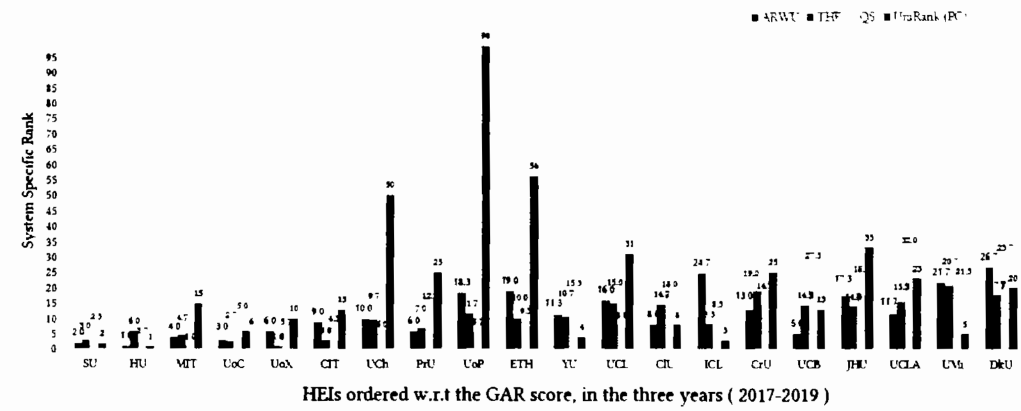


Figure 4.6 Top 20 HEIs w.r.t the GAR score and their Research Productivity

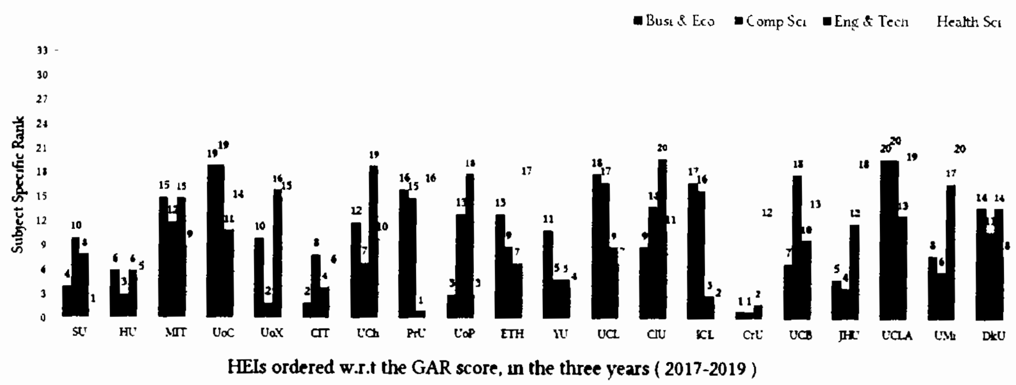


Figure 4.7 Subject-Specific Ranks of top 20 HEIs, in years 2017—2019

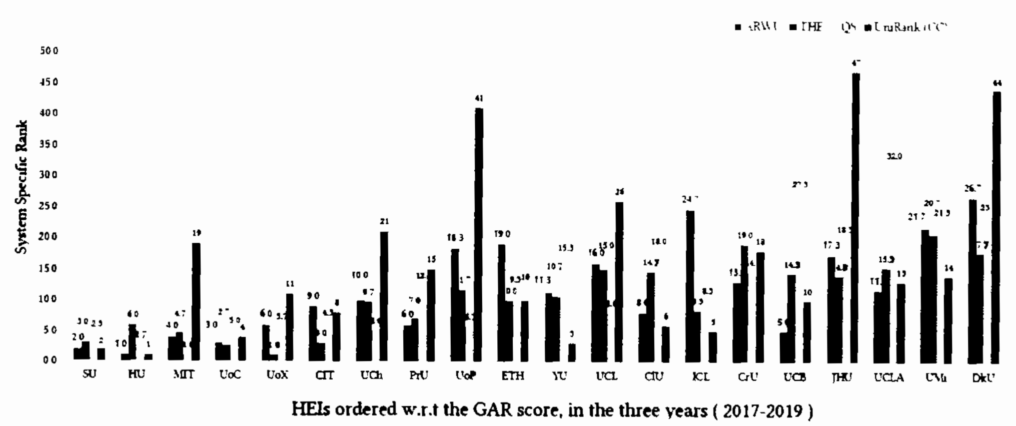


Figure 4.8 Top 20 HEIs w.r.t the GAR score in years 2017—19, and Influential Research Faculty

4.6.4. Analysis of the Influential Research Faculty

In experts' opinions [69, 2], 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. These ranking dimensions were also analyzed. The affiliations of the influential researchers (in all the disciplines) were mapped with the top-performing HEIs according to the GAR score. A part of the mapping results is presented in Figure 4.8. 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.

4.6.5. Considering the context in the Ranking

One of our data sources was the DBpedia data repository. Various meaningful information about the countries and HEIs were extracted; including the base country of an HEI, its exact location, city, year of establishment, etc. The availability of the year of establishment of HEIs helped us in assessing their academic age. A part of the result is presented in Figure 4.9. An overview of the graph reveals the fact that all the top 20 HEIs with outstanding performance have academic age of more than a hundred years. It means achieving outstanding status at the world level, which requires decades of consistent effort.

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, the relationship between the HDI of a country and the count of its top-performing HEIs was explored. The results of the analysis are presented in Figure 4.10. 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. The analysis 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 4.11 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, the countries having a considerable number of HEIs in their different cities, were also identified. Quite understandably, the top fifteen countries having HEIs with outstanding performance, are the most stable and prosperous courtiers. Figure 4.12 represents a part of the analysis. Governments of developed countries are cognizant of the role of HEIs in the advancement of their socioeconomic 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 [35], the argument was discussed, and the comparisons of HEIs having very different sizes, budgets, 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 a 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 the such fine-grained analysis.

4.7. Summary

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 researchers' ranking methodology, the affiliations of the influential researchers were mapped with the top-performing HEIs, to explore their research productivity and research impact. The effectiveness of the proposed data analysis strategy was demonstrated through the publicly verifiable data repositories – the DBpedia and Google Scholar Citations. The research work reported in this chapter is a continuation of the previous research effort which proposed the OpenRank academic ranking methodology.

Table 4.4. Top 20 HEIs w.r.t the Influential Research Faculty (overall) as of Dec 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

Table 4.5. 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)	285598 (382869)	364.93
Anil K. Jain	Michigan State University	182 (198)	995 (1017)	192561 (225588)	393.76
Jiawei Han	University of Illinois Urbana-Champaign	168 (183)	1446 (1564)	170320 (157743)	488.09
Federico Calzolari	Scuola Normale INFN CERN	120 (137)	2722 (2841)	83485 (107508)	2637.54
Thomas S. Huang	University of Illinois, Urbana-Champaign	143 (165)	1993 (2092)	150248 (176152)	718.22

* Current statistics (June 2021) are inside the brackets

Table 4.6. Top 20 HEIs w.r.t the Research Publications by their IRF, as of Nov 2019.

Rank	Higher Education Institution (HEI)	Acronym	Total Publications	UniRank Score
1	Stanford University	SU	25298	2
2	Harvard University	HU	53598	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	13444	4
12	University College London	UCL	2952	31
13	Columbia University	ClU	7999	8
14	Imperial College of Science Technology and Medicine London	ICL	17238	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	12114	5
20	Duke University	DkU	4196	20

Table 4.7. 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)	479415 (514877)	1699.80
Robert Kaplan	Harvard University	94 (100)	657 (833)	165019 (192672)	4079.41
Lawrence Summers	Harvard University	174 (184)	1523 (1668)	142894 (166098)	1855.90
Kenneth J. Arrow	Stanford University	149 (162)	1857 (2037)	205116 (237434)	1256.44
James Heckman	University of Chicago	164 (173)	1063 (1158)	190164 (228053)	571.73

* Current statistics (June 2021) are inside the brackets

Table 4.8. Top 20 HEIs with their rank in the different disciplines w.r.t the IRF, as of Nov 2019.

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	16 (3)	1 (26)	2 (21)	0 (n.a)	12.78
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 4.9. Top 5 Researchers in Health Sciences w.r.t the UniRank methodology (Nov 2019)

Researcher	Affiliation	H-Index	Publication Count	Citation Count	APP Score
Frank B. Hu	Harvard University	234 (278)*	2001 (2363)	226525 (360114)	475.84
Dr. Joann E. Manson	Harvard University	266 (298)	1500 (2822)	273626 (369557)	346.04
Daniel Levy	National Heart Lung and Blood Institute	207 (245)	1976 (2101)	213572 (319635)	535.98
Guido Kroemer	Université de Paris Hôpital Européen	227 (247)	1597 (1885)	225593 (282207)	425.75
Harlan Krumhold	Yale University	191 (214)	2001 (2417)	177379 (252018)	511.20

Table 4.10. Top 20 HEIs w.r.t the Research Impact by their IRF, as of Nov 2019.

Rank	Higher Education Institution (HEI)	Acronym	Citation Count	UniRank
1	Stanford University	SU	3282264	2
2	Harvard University	HU	7740062	1
3	Massachusetts Institute of Technology	MIT	327980	21
4	University of Cambridge	UoC	1190102	6
5	University of Oxford	UoX	579673	11
6	California Institute of Technology	CIT	16602	157
7	University of Chicago	UCh	285775	27
8	Princeton University	PrU	373859	19
9	University of Pennsylvania	UoP	161276	50
10	Swiss Federal Institute of Technology (ETH) Zurich	ETH	0	n.a
11	Yale University	YU	1529522	3
12	University College London	UCL	294022	26
13	Columbia University	ClU	1209813	5
14	Imperial College of Science Technology and Medicine London	ICL	1024013	7
15	Cornell University	CrU	310595	23
16	University of California at Berkeley	UCB	788457	10
17	Johns Hopkins University	JHU	231683	35
18	The University of California at Los Angeles	UCLA	511565	13
19	University of Michigan	UMi	416320	16
20	Duke University	DkU	179746	44

Table 4.11. 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)	130168 (159570)	1040.41
Yang	University of California Los Angeles	149 (161)	3000 (3000+)	101873 (122185)	816.65
Sercan Sen	Hacettepe University	164 (187)	2101 (2785)	125329 (159671)	572.37
Lev Dudko	Lomonosov Moscow State University	164 (186)	1571 (1963)	144708 (188347)	399.32
Rob Knight	University of California San Diego	166 (199)	675 (890)	167776 (257855)	146.82

* Current statistics (June 2021) are inside the brackets

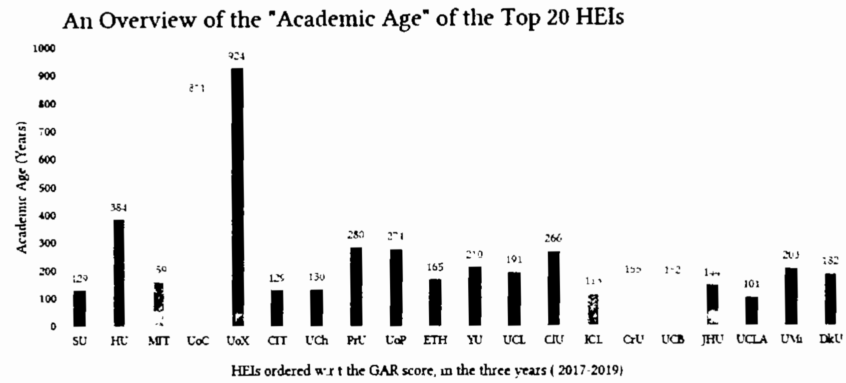


Figure 4.9. An overview of the "Academic Age" of top-performing HEIs

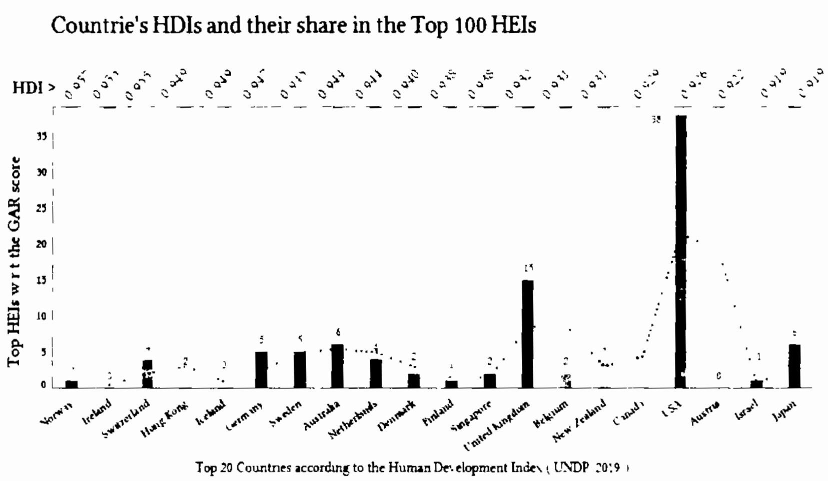


Figure 4.10. Top 20 Countries with the HDI (UNDP, 2019) vs top HEIs (GAR Score)

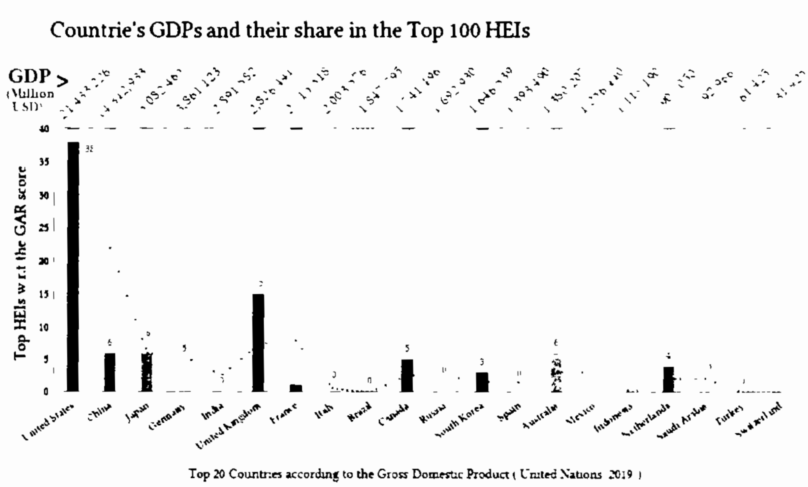


Figure 4.11. Top 20 Countries with the GDP Per Capita (UN, in 2019) Vs top HEIs (w.r.t the GAR Score)

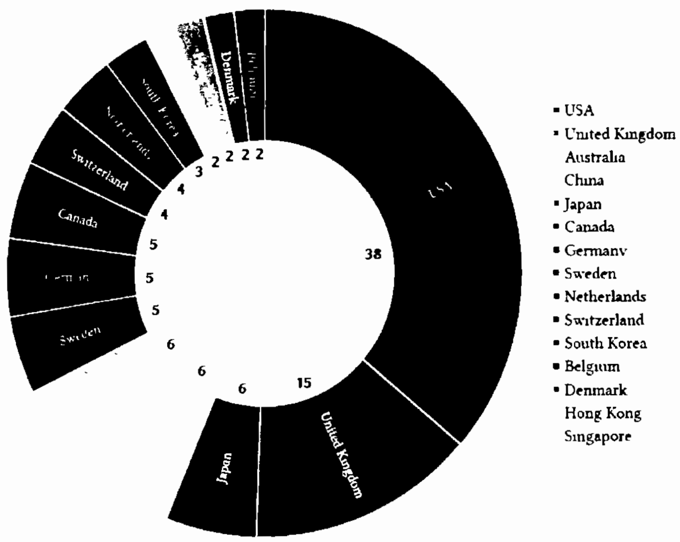


Figure 4.12. Top 15 Countries having the top perfuming HEIs w.r.t the GAR score.

Using the proposed researchers' ranking methodology (ResRank), the eight hundred researchers belonging to the four scientific disciplines (Computer Sciences, Engineering, Health Sciences, and Economic & Management Sciences) were ranked. Then, the list of the top hundred global HEIs was produced. These HEIs had unanimously declared outstanding performance by the three reputed academic ranking systems. Finally, the affiliations of the influential researchers were mapped with the top-performing HEIs, to explore their research productivity and research impact.

The analysis interprets the ranking results according to various scientific disciplines and ranking dimensions such as *Academic Quality*, *Research Productivity*, *Graduate-Employability*, *Industry-Academia Linkage,* and *Internationalization.* The analysis highlighted various aspects of the HEIs' performance for the end-user of the rankings.

The use of the objective indicators extracted from two well-known publicly verifiable data repositories makes the ranking results reproducible and increases the confidence of the academic stakeholders. Moreover, diverse data sources provide a comprehensive understanding of the factors that catalyze higher education. The resultant academic rankings make the ranking process more transparent and more representative, thereby the ranking results less controversial.

Chapter 5

Heading Towards the Sub-Discipline Specific Rankings

Although the academic rankings of Higher Education Institutions(HEIs) face considerable criticism, they are here to stay. Having become competent in assigning holistic ranking scores to HEIs, reputed ranking entities have now started focusing on subject-specific and regional rankings. However, in experts' opinion, the process of assigning rankings should be more consistent, transparent, and representative. This study focuses on enhancing the credibility of the academic ranking process, by performing a fine-grained assessment of the academic data about the Computing discipline. The proposed assessment approach explores the data at the sub-discipline level, analyzing several ranking dimensions including the research productivity, research impact, and research contribution of influential research scholars affiliated with renowned HEIs in the Computing discipline.

The analysis considers highly curated data published by three well-known international academic ranking entities, namely, ARWU, THE, and QS, in 2018, 2019, and 2020. Researchers' profiles are obtained from the Scopus repository, and the DBpedia repository is used to retrieve information about HEIs and their locations. For a stable comparison of the subject-specific academic rankings, the Grand Average Rank measure is employed, whereas for finding the most influential researchers in Computing, the ResRank measure is used. The sub-discipline-specific academic rankings provide more detailed insight into the academic rankings, thereby providing more robust decision support. This analysis, which focuses on the Computing sub-discipline, is among the first few such efforts.

5.1. Introduction

Throughout the world, Higher Educational Institutions (HEIs) catalyze the human development process at the local and global levels. They enhance the competitiveness of a nation and promote its development by fostering research and hi-tech innovation. In developed countries, HEIs focus on nurturing innovation and commercialization. HEIs use systematic ways to assess their academic performance [70]. They are regularly assessed by various public and private ranking entities such as the Academic Rankings of World Universities (ARWU), Quacquarelli Symonds (QS), and Times Higher Education (THE). These ranking entities have been publishing academic rankings since 2003 [50].

In the last few years, the academic ranking process has been steadily evolving, and ranking results are enabling stakeholders to take well-informed decisions. However, academic experts and users have some reservations regarding HEI ranking results. The validity and transparency of the rankings have been questioned, as highlighted in previous studies [37, 4, 14, 54]. Among the prominent issues are *controversial ranking methodologies*, the *use of subjective ranking indicators*, the *validity of the ranking results*, *transparency of the data sources* [1], *misinterpretation of academic performance*, and *limited coverage of global HEIs* [14].

Usually, ranking entities rely on an arbitrary combination of ranking indicators and provide a single numeric value as the ranking score of an HEI. These indicators include the reputation of a certain university amongst its peers, learning environment (such as student-to-teacher ratio and graduate-employability), research quality, funding, and student acceptance and graduation rates [50, 71, 35]. However, a single numeric value usually hides the impact of individual KPIs. The administrative stakeholders of HEIs need insight into these indicators to select the specific areas that require strategic investment. Consequently, realizing the need for more fine-grained analysis, various reputed ranking systems, ARWU[10], THE[11] , and QS[12], have started publishing the Subject-Specific ranking of HEIs [2].

The analysis focused on the three-year (2018, 2019, and 2020) subject-specific ranking scores for the Computing discipline allocated to renowned global HEIs by THE, ARWU, and QS. To obtain a normalized ranking score, the three-year average of the ranking scores was calculated for each of the HEIs. For descriptive analysis, the average of the nine rankings was computed using the ranking lists produced by the three ranking entities. The statistics reveal significant differences in the ranking outcome by the three ranking entities, even within a single discipline (Computing), for similar HEIs.

---

[10] http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/index.html [Accessed on August 2021]
[11] https://www.timeshighereducation.com/world-university-rankings/2021/world-ranking [Accessed on July 2021]
[12] https://www.topuniversities.com/subject-rankings/2020 [Accessed on May 2021]

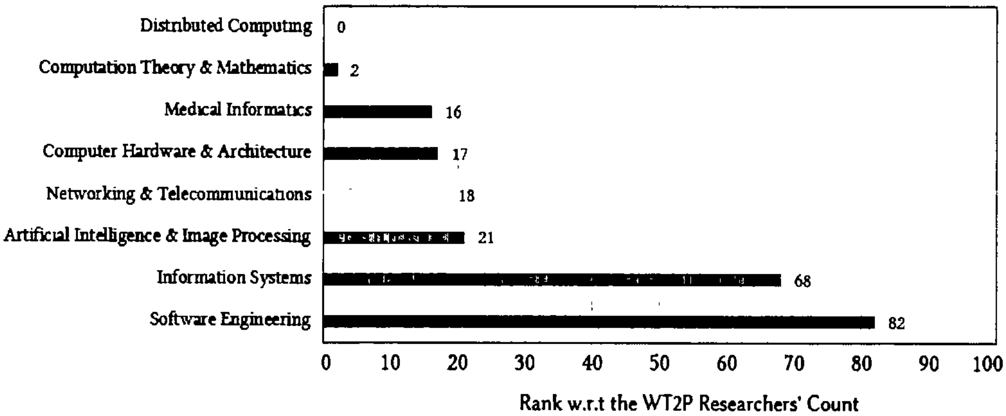


Figure 5.1. MIT positions in the eight Sub-Disciplines of Computing

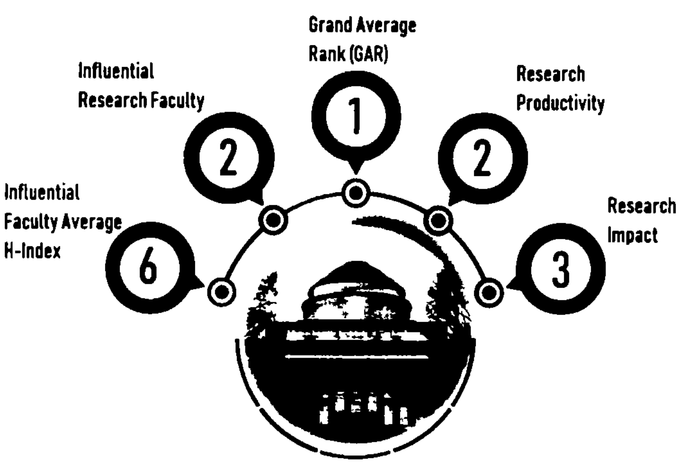


Figure 5.2. MIT positions w.r.t the various ranking dimensions or indicators

A summary of the analysis is presented in Table 4.1 and a part of the results is visualized in Figure 5.3 and Figure 5.5. For instance, the Massachusetts Institute of Technology (MIT) appears in the first five positions of the subject-specific rankings (Computing) published by THE, ARWU, and QS in the years 2018, 2019, and 2020. Therefore, it attains the first rank as per the GAR score. The GAR score is an overall average of the nine raking scores published by ARWU, THE, and QS for an HEI in 2018, 2019, and 2020. Nevertheless, there is considerable variation in ranks assigned to the eight sub-disciplines of Computing within MIT, as shown in Figure 5.1.

The *ranking-dimension-oriented* analysis of the HEIs' data recently published by a group of researchers at Stanford University, USA, must be mentioned here. Based on their proposed researchers' ranking metric, the "*composite indicator,*" they published profiles of the "World's top 2-percent researchers" (150,000+) in various scientific disciplines [24]. A detailed discussion of the dataset is given in e). The dataset was exploited to analyze the subject-specific rankings (computer science) of the global HEIs published by ARWU, THE, and QS in the stipulated three years. Using the dataset, the subject-specific rankings are analyzed in terms of ranking indicators such as the *Significant Research-Contribution*, *Research-Impact,* and *Research-Productivity* of the influential researchers serving at the top-ranked global HEIs. The analysis also provides better insight into the subject-specific rankings of HEIs in the Computing discipline. The outcome of the analysis for MIT is presented in Figure 5.2. For more detail, please consult Table 5.4.

The results of the analysis present different pictures of even the top 10 HEIs in the computer science discipline. Such inconsistencies provide supportive arguments to the critics of the academic ranking process. These differences need to be addressed. Nevertheless, finding suitable solutions for all the issues requires considering time and resources. A research effort has already been done, focusing on enhancing the credibility of the ranking process by ranking the global HEIs using objective ranking indicators that were extracted from publicly verifiable data sources. This work is reported in [35]. The OpenRank methodology suggested in the research was developed for the holistic ranking of the HEIs. Nevertheless, the need for a fine-grained analysis at the subject-specific level was felt seriously. In another research effort, the same research objectives were achieved using the *ResRank* methodology. This work focused on enhancing the *validity* and *transparency* of the ranking process. It also performed an in-depth assessment of the rankings in various sub-disciplines of the Computing discipline. The research contributions are highlighted here:

- An in-depth analysis of the academic ranking statistics
- Producing the sub-discipline-specific rankings of HEIs in the Computing discipline
- Highlighting the need for sub-discipline-specific academic rankings in the future
- Demonstrating fine-grained data analysis based on public data repositories

The rest of the paper is organized as follows: Sections 5.3 and 5.4 articulates the research problem, Section 5.5 describes the fine-grained analysis approach, Section 5.6 provides details of the experiments and their results, and Section 5.7 provides a discussion of the results. Finally, Section 5.8 provides the conclusion and highlights research directions for the future.

5.2. The Benchmark Rankings Systems and Data Sources

A brief description of three reputed ranking entities is given below, to provide a thorough understanding of the proposed research methodology, which will be beneficial for better analysis. The process of academic rankings usually involves a complex ranking methodology that is realized in terms of ranking indicators or parameters. The indicators reflect the aims and objectives of the ranking entity. They determine the effectiveness of a ranking system and differentiate it from others. An effective academic ranking methodology:

a) Employs comprehensive parameters
b) Prefers quantitative measures over qualitative measures
c) Uses the publicly verifiable data
d) Enables the users to "Drill Down and Roll Up" the indicator's detail.

Rigorous ranking methodologies prefer employing objective indicators and publicly verifiable data sources. Scopus, DBpedia, GSC, and ArnetMiner repositories are examples of such data sources [57]

5.2.1. Some Well-known HEIs Ranking Systems

Multiple ranking systems are gauging and publishing the academic performance of HEIs worldwide. However, the three reputed academic ranking systems – ARWU[13], THE[14], and QS[15]. 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 [37, 35].

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

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

[15] https://www.topuniversities.com/subject-rankings/methodology [Accessed on Oct 2021]

Therefore, their introduction is abridged and the subject-specific ranking methodologies are focused. The ranking methodologies employed by the three well-known academic ranking systems have been introduced in Section 1.7 and various issues of the ranking process were also highlighted in Section 1.4. In this research work, the analysis is performed at the level of ranking indicators i.e., the research productivity, research impact, and research contribution of the influential researchers affiliated with the top-ranked HEIs in the Computing discipline. The association of the most significant researchers ("World's top-2 percent researchers") was mapped with the world's top 100 HEIs in the Computing domain, as reported by ARWU, QS, and THE in the stipulated three years. All the researchers' profiles are indexed in the Scopus repository. The mapping helped us analyze the data related to influential researchers working with the top-ranked HEIs in the Computing domain. The analysis helped in producing the subdiscipline-specific HEI rankings.

5.2.2. The Analysis Approach and Data Sources

The sub-discipline-specific ranking process comprises three phases. First, a list comprising of HEIs was produced. They were having top performance in the years 2018, 2019, and 2020 using the subject-specific rankings provided by the three ranking entities. Second, for an in-depth analysis, the Scopus repository was employed to extract the profiles of the influential researchers in various sub-disciplines of Computing. Finally, the affiliations of the influential researchers were determined and mapped to the top-ranked HEIs in the Computing discipline. Before a detailed description of the three phases, the data sources and evaluation metrics are described below. The analysis is based on the highly curated data produced and published by the three well-known international academic rankings – ARWU[16], THE[17], and QS[18]. The researchers' profiles were obtained from a derivative dataset of the Scopus repository as introduced in Section e), while the DBpedia repository

---

[16] http://www.shanghairanking.com/ARWU-Methodology-2020.html [Accessed on Dec 2020]
[17] https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2020-methodology
[18] https://www.topuniversities com/subject-rankings/methodology

was employed for the information about HEIs and countries. The dataset is introduced in Section a).

The subject-specific (Computer Science) rankings of HEIs were produced by processing the ranking results published in the last three years 2018—2020 by the three reputed academic ranking systems. As per the research objective, for fine-grained analysis, reliable and publicly verifiable data were used. These datasets were produced by well-known ranking systems and can be verified on their official web portals. Some of the systems also provide a limited amount of data in spreadsheets or flat files.

Heading toward the fine-grained analysis of the academic performance evaluation data, a systematic approach was adopted. Initially, we generated a list of these HEIs with top performance in the years 2018, 2019, and 2020 using the ranking results published by the three reputed 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, the list of HEIs was determined, which repeatedly maintained their position in the list of top-100, for three consecutive years. The 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, the three lists were combined to produce the final list of top-performing HEIs in the stipulated three years.

After finding the top 100 HEIs, the 3-year average of the ranking scores for an HEI was calculated. The average was computed for each of the three ranking systems to obtain a stable ranking score. A comparison of the 3-year average scores also highlighted the significant differences among the ranking systems even for a single HEI. The overall average (Grand Average) of the nine ranking lists for each of the HEIs was computed also. It was based on the rankings published by the three systems in the three years. The mathematical model for the selection of the top hundred HEIs is given below:

$$\text{Top100HEIs} = \bigcup \left(\forall\ \text{RankEnt} \bigcap_{Y=2018}^{2020} \text{Top100HEIs}_{\text{RankEnt}} \right) \quad (1)$$

Table 5.1 Researchers' information extracted from the Scopus repository.

Attribute	Acronym	Description
Research Productivity	PC	Number of the total research publications by a researcher
Research Impact	CC	Number of the total citations (including the self-citations) received by research publications of a researcher
	XSCC	Number of the total citations (excluding the self-citations) received by research publications of a researcher
Significant Research Contribution	SRC	Number of research publications by a researcher as the first author or as the single author
H-Index	HI	Value of the H-Index of a researcher, as recorded at the end of 2019.
FirstYear	FY	The year of first publication by a researcher
LastYear	LY	The year of most-recent publication by a researcher
Research Domain	RD	Name(s) of the scientific research domain of the author
Research Area	RA	Name(s) of the research area(s) in the research domain (RD)
Author's Affiliation	Affiliation	Name of the research institution, with which the author is currently affiliated
Authors' Country	Country	Name of the base country of the author

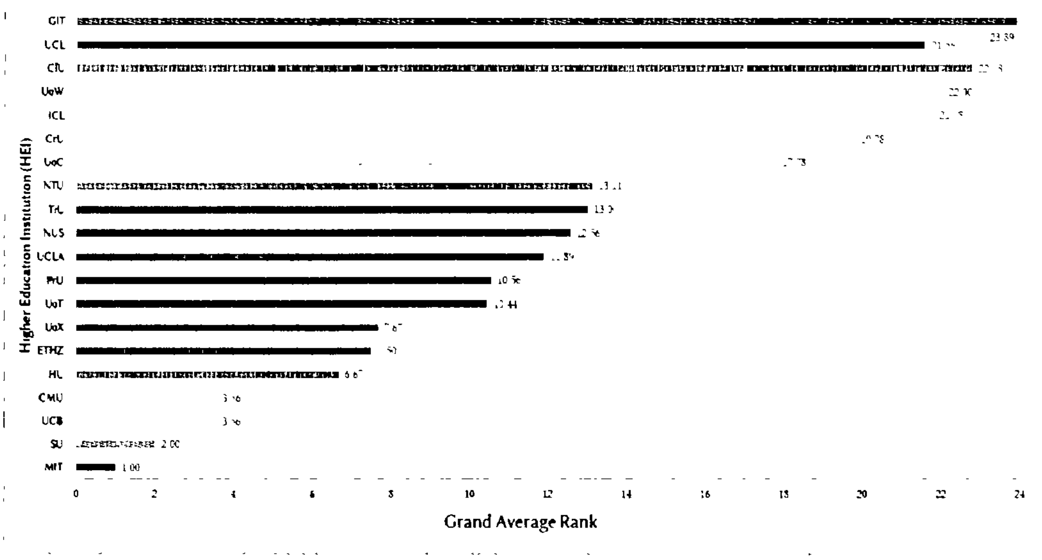


Figure 5.3. Top 20 HEIs w.r.t the Average Rank in Years 2018, 2019, and 2020.

where RankEnt ϵ { ARWU, THE, and QS }

The names and scores of the top-20 HEIs w.r.t the Grand Average Rank (GAR) top-20 HEIs are presented in Table 5.2. For a quick overview, the data is visualized in Figure 5.3. According to the GAR, Harvard University (HU) leads the top-performing HEIs with a score of 1.0. whereas the Georgia Institute of Technology (GIT) has a score of 23.89, in the twentieth position.

The profiles of the influential research scholars were extracted from the Scopus repository. Scopus maintains various valuable statistics about thousands of researchers and millions of their research publications. For the research analysis, various statistics were extracted from the repository including the Research Productivity (or Publication Count – PC), Research Impact (or Citation Count – CC), H-Index (HI), and Author's Significant Research Contribution (SRC). The SRC is the count of the research publications by a researcher as the single or as the first author. A brief description of the selected attributes is presented in Table 5.1. and the motivation behind the selection is summarized here.

The selection of the first three parameters is based on two facts. First, in experts' opinions, these parameters (PC and CC) are treated as the Key Performance Indicators (KPIs) in the related literature [15, 18]. Second, most of the bibliographic data repositories essentially record them. The meta-data of Scopus, Web of Science, GSC, DBLP, Microsoft Academics, and the ArnetMiner data repositories, is evident in the fact [21]. The last parameter SRC is specific to the Scopus repository as it is difficult to find the first year of publication of an author in other data repositories including the GSC and MAG. Nevertheless, the importance of the parameter can easily be understood in terms of the significant research contribution of a researcher.

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 [59]. It is the count of the citations of a researcher's 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 and quality of work of a research scholar [61].

Although the *Hirsh-Index* faces many critics, nevertheless, the wisdom behind the measure is very foundational; therefore, the metric in its original, or its variants has been surviving for the last fifteen years, in the bibliographic data repositories [62, 63]. Finally, the indicator *Significant Research Contribution* (SRC) score is discussed. The SRC score is an effort to gauge the leading research contribution of a research scholar. 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. Such measures also have multiple variants in the literature [38, 64]. Nevertheless, since our research work is focused on mapping affiliations of the influential researchers with the top-performing HEIs; therefore, to gauge the significant contribution of a researcher, the count of the research publications by him or her was considered, as the single or first author.

After the selection of the indicators, their weights were to be decided. The weights used for the ranking parameters show their importance in the process of performance evaluation. It seems suitable to mention that there is no standard weighting criterion for the selected ranking indicators, therefore their weights usually remain controversial. The well-known InfoGain (IG) [32] algorithm was applied to the target dataset, to determine the weights (w_1, w_2, w_3, *and* w_4) of the indicators.

InfoGain algorithm requires a test dataset with a class label; the "c-score" introduced in [43] was employed as the default criterion to rank the researchers to obtain the test dataset with the binary class label "highly Influential" or "regular". Equation 2 represents the process of applying InfoGain to the target dataset (*R)* along with the class label (*C*).

$$\text{InfoGain}(\text{R}, C) \rightarrow \quad w_1\text{HI},\ w_2\text{CC},\ w_3\text{PC},\ w_4\,\text{SRC} \qquad (2)$$

The InfoGain algorithm, produced $w_1 = 0.41$, $w_2 = 0.38$, $w_3 = 0.09$, and $w_4 = 0.12$ as the weights for the indicators. Equation 3 represents the researchers' ranking methodology with the weighted parameters. It is worth mentioning that GSC orders the researchers' profiles for their *Research Impact* or *Citation Count*. Nevertheless, agreeing with the fact that alone quantity does not ensure quality, a more comprehensive methodology was adopted to rank the research scholars. After completing the pre-requisite calculations, the methodology can be expressed as follows:

$$ResRank\ (RP) = 0.41\mathrm{HI} + 0.38\mathrm{CC} + 0.09\mathrm{PC} + 0.12\ \mathrm{SRC} \qquad \textbf{(3)}$$

Where the *ResRank ( )* is the function to rank a researcher's profile (*RP*) as influential or regular. 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 Equation 3. The dataset produced using Equation 3, is the subject-specific ranking of the research scholars belonging to the Computer Science discipline. Earlier, we also developed a list of the HEIs that consistently retained their position in the top 100 HEIs with outstanding performance, in the stipulated three years.

After having the required statistics for the fine-grained analysis of the academic rankings for *Research Productivity* the two datasets were mapped – *Influential Researchers* and *TopHEIs* and ranks the top-performing HEIs w.r.t the count of the influential researchers. The mapping highlighted the HEIs having the highest number of influential researchers. The fine-grained analysis proceeded with the same pattern and developed the HEIs rankings for the count of their *Research Publications* (PC) and *Research Impact* or *Citation Count* (CC). The analysis provides insight into the end-users of the HEIs rankings. The results of the analysis are presented in section 4.6.

5.3. Experiments

The required data were extracted from the data sources introduced in Section 4.3, using multiple techniques. To generate the list of top-100 HEIs with consistently outstanding performance in the stipulated three years, the relevant data was retrieved from official web portals of ARWU, THE, and QS academic ranking systems. The profiles of influential researchers were extracted from the publicly available derivative [43] of the Scopus data repository. To obtain the necessary data about the HEIs and countries, the DBpedia data repository was employed.

Table 5.2. HEIs' Average Rank (Computer Science) by the three ranking systems and GAR Score

	HEI	Acronym	ARWU			THE			QS			3-Year Rank Average			GAR Score
			2018	2019	2020	2018	2019	2020	2018	2019	2020	ARWU	THE	QS	
1	Massachusetts Institute of Technology	MIT	1	1	1	2	5	4	1	1	1	1.00	1.00	1.00	1.0
2	Stanford University	SU	2	2	2	1	3	2	2	2	2	2.00	2.00	2.00	2.0
3	University of California at Berkeley	UCB	5	3	3	4	4	4	4	4	4	3.67	4.00	3.00	3.6
4	Carnegie Mellon University	CMU	3	4	4	6	6	6	3	3	3	3.67	3.00	4.00	3.6
5	Harvard University	HU	8	6	6	11	9	8	6	7	7	6.67	6.67	6.67	6.7
6	Swiss Federal Institute of Technology (ETH) Zurich	ETH Z	4	5	5	4	2	3	9	9	9	4.50	9.00	9.00	7.5
7	University of Oxford	UoX	13	10	10	3	1	1	7	6	5	11.00	6.00	6.00	7.7
8	University of Toronto	UoT	11	12	9	22	18	23	10	11	10	10.67	10.33	10.33	10.4
9	Princeton University	PrU	10	9	12	12	8	9	8	13=	11	10.33	10.67	10.67	10.6
10	The University of California at Los Angeles	UCL A	6	8	11	13	13=	15	13	13=	15	8.33	13.67	13.67	11.9
11	National University of Singapore	NUS	15	16	18	13	15	11	10=	10	12	16.33	10.67	10.67	12.6
12	Tsinghua University	TiU	7	7	7	20	20	15	20	15=	13	7.00	16.00	16.00	13.0
13	Nanyang Technological University	NTU	9	13	8	31	29	13	16	12	16	10.00	14.67	14.67	13.1
14	University of Cambridge	UoC	51-75	27	38	5	4	5	5	5	6	42.67	5.33	5.33	17.8
15	Cornell University	CrU	12	14	14	15	14	14	25	25	19=	13.33	23.00	23.00	19.8
16	Imperial College of Science Technology and Medicine London	ICL	40	40	34	9	11	7	12	15=	14	38.00	13.67	13.67	21.8
17	University of Washington	UoW	33	26	33	17	22	26	18=	17	18	30.67	17.67	17.67	22.0
18	Columbia University	ClU	25	22	36	19	20	21	24	18=	19=	27.67	20.33	20.33	22.8
19	University College London	UCL	20	19	23	18	23	19	32	23	17	20.67	20.00	24.00	21.6
20	Georgia Institute of Technology	GIT	24	18	29	8	7	16	21	24	27	23.67	24.00	24.00	23.9

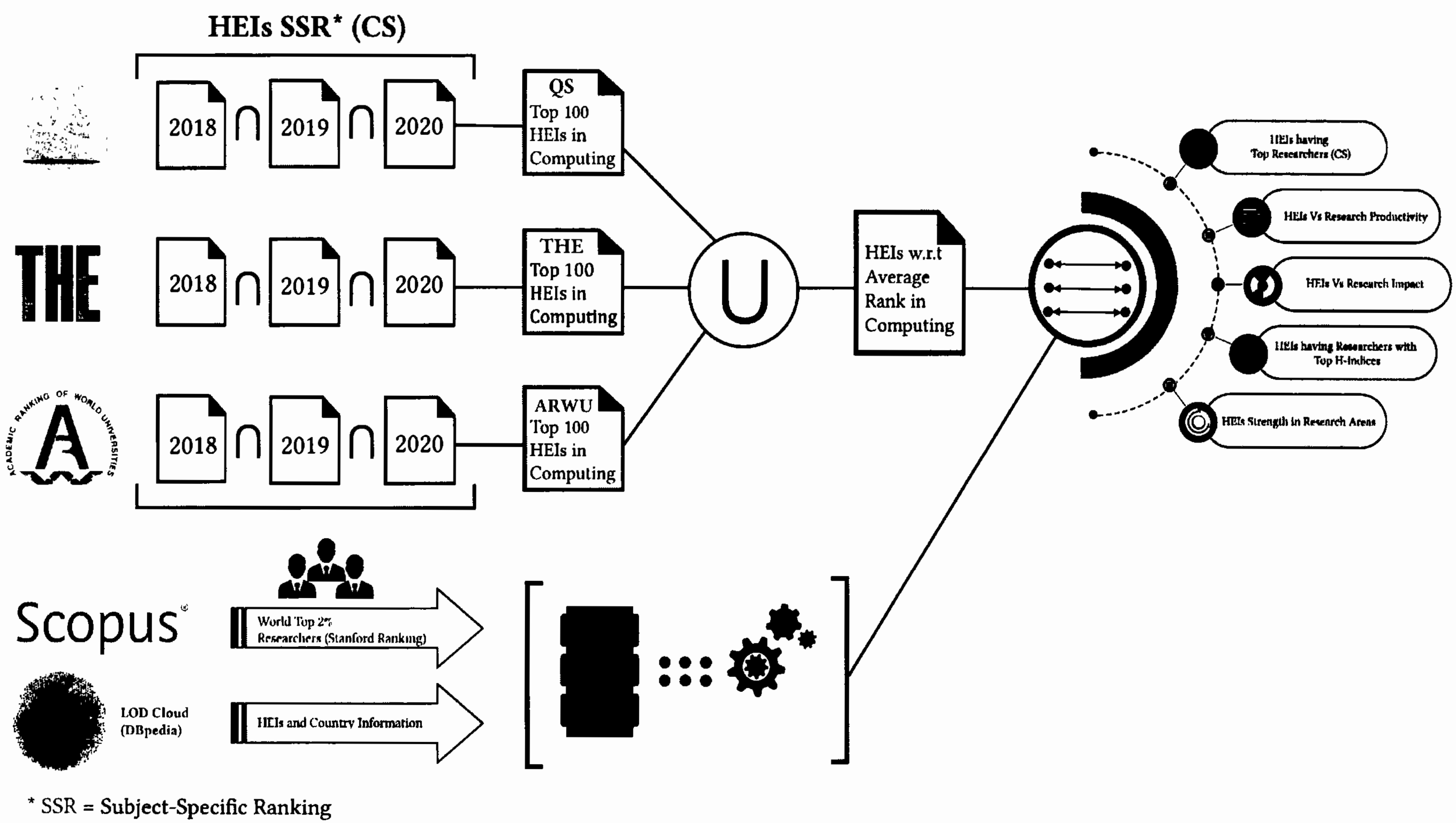


Figure 5.4. Workflow of the Fine-Grained HEIs Ranking Strategy

Table 5.3 HEIs' (Computer Science) Rankings w.r.t various Ranking Indicators

	Higher Education Institution (HEIs)		**GAR** Score	**RF (CS)** Count & Rank		**PC** Count & Rank		**CC** Count & Rank		**XSCC** Count & Rank		**AHI** Count & Rank	
1	Massachusetts Institute of Technology	MIT	**1.0**	81	2	12936	**2**	788524	3	788112	**3**	37.56	6
2	Stanford University	SU	**2.0**	78	3	12685	**3**	953822	1	953451	**1**	40.50	3
3	University of California at Berkeley	UCB	**3.6**	65	4	10374	**6**	923048	2	922729	**2**	41.97	1
4	Carnegie Mellon University	CMU	**3.6**	95	1	15318	**1**	639417	4	638905	**4**	34.74	10
5	Harvard University	HU	**6.7**	22	20	3087	**20**	148098	19	148000	**19**	33 27	14
6	Swiss Federal Institute of Technology (ETH) Zurich	ETHZ	7.5	42	10	9719	**9**	409014	6	408652	**6**	40.93	2
7	University of Oxford	UoX	7.7	34	16	5817	**15**	289029	13	288797	**13**	34.65	11
8	University of Toronto	UoT	**10.4**	56	6	9425	**10**	445733	5	445465	**5**	32.89	15
9	Princeton University	PrU	**10.6**	39	13	8151	**11**	345861	8	345619	**8**	38.15	5
10	The University of California at Los Angeles	UCLA	**11.9**	44	9	11669	**5**	403585	7	403254	**7**	38.80	4
11	National University of Singapore	NUS	**12.6**	45	8	9838	**7**	263593	14	263346	**14**	33.73	13
12	Tsinghua University	TiU	**13.0**	40	11	12189	**4**	246915	15	246629	**15**	37.30	7
13	Nanyang Technological University	NTU	**13.1**	27	19	5339	**17**	164054	18	163907	**18**	32.07	17
14	University of Cambridge	UoC	**17.8**	33	17	4400	**19**	184680	17	184562	**17**	29.15	20
15	Cornell University	CrU	**19.8**	40	12	5526	**16**	294733	11	294594	**11**	33.80	12
16	Imperial College of Science Technology and Medicine London	ICL	**21.8**	36	14	7318	**12**	227494	16	227181	**16**	31.58	18
17	University of Washington	UoW	**22.0**	50	7	6295	**14**	315876	10	315686	**10**	35.02	9
18	Columbia University	ClU	**22.8**	36	15	6764	**13**	332235	9	332044	**9**	36.47	8
19	University College London	UCL	**21.6**	32	18	4502	**18**	142752	20	142600	**20**	29.72	19
20	Georgia Institute of Technology	GIT	**23.9**	57	5	9779	**8**	289185	12	288899	**12**	32.44	16

PC = Publication Count, **CC** = Citation Count, **XSCC** = (CC without self-citations), **RF(CS)** = Research Faculty (CS), **AHI** = Average H-Index

The data in the DBpedia was processed through its SPARQL Endpoint[19] using the SPARQL queries [58]. 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.).

Data processing was with the selection of the top-100 HEIs with outstanding performance in the stipulated three years (2018, 2019, 2020) from the HEIs' rankings published by the ARWU, THE, and QS. The selection method is modeled in Equation 1. After obtaining the list, the HEIs were sorted according to the average of their scores in nine rankings published by the three systems in the three years. Qualifying the ranking criteria of the three rankings systems and retaining position in their lists of top-100s, advocates outstanding performance and HEI. The researchers' ranking methodology was applied to the dataset containing profiles of the researchers, for evaluation of their influential status. The mathematical expression of the methodology is presented in Equation 2.

Finally, the researchers' affiliations were mapped with the top-performing HEIs in the Computer Science discipline, to obtain a fine-grained ranking of HEIs w.r.t the Computing research faculty. The list of the top twenty HEIs w.r.t their research Faculty is presented in Figure 5.6. The HEIs were also ranked according to the count of their research publications (PC) published by the researchers affiliated with them. The list of the top twenty HEIs, w.r.t the research publications is presented in Figure 5.7.

Gauging the impact of the research publication was also part of the fine-grained analysis as suggested in [28]; therefore, the HEIs were ranked according to their research citation count (CC) earned by the researchers affiliated with them. The list of the top twenty HEIs, w.r.t the citation count, is presented in Figure 5.8. The Scopus data repository maintains the count of the research citations excluding the self-citations of a researcher. This count represents the significant research contribution of a researcher. Therefore, the HEIs were ranked according to the research impact (excluding self-citations). The result of this ranking is visualized in Figure 5.10.

---

[19] https://dbpedia.org/sparql [Accessed on December 2021]

The HEIs were also ranked according to their significant researchers with outstanding research contributions. The result of this ranking is presented in Figure 5.11. The statistics related to the above-mentioned analysis are summarized in Figure 5.10. The DBpedia repository also provided the dates of establishment of the HEIs. These statistics were used to determine the age of the HEIs. The ages of the twenty HEIs having outstanding performance, are presented in Figure 5.10.

5.4. Results and Discussion

Before starting a discussion on the results, it is worth mentioning the issue of data incompleteness, for the intended analysis. Although, the academic data available on the Internet is abundant and perpetually increasing. Nevertheless, most of the data repositories, including DBpedia and Scopus, are not purposely built for academic rankings. These data sources were exploited for the information on the global HEIs having a presence in these repositories.

5.4.1. Analysis of the Deviation in the Rankings

The Grand Average Rank (GAR) was calculated for the top-ranked HEIs in the Computer Science discipline. GAR is the average of the nine ranking scores, allocated to the HEIs by ARWU, THE, and QS ranking systems. Based on the GAR score, the list of the top 100 HEIs was compiled. These HEIs had consistent performance according to the ranking methodologies. GAR score is a rigorous indicator of the top-performing HEIs, as qualifying the ranking criteria of the three ranking systems and retaining a position in their lists of top-100s, is hard enough for an HEI.

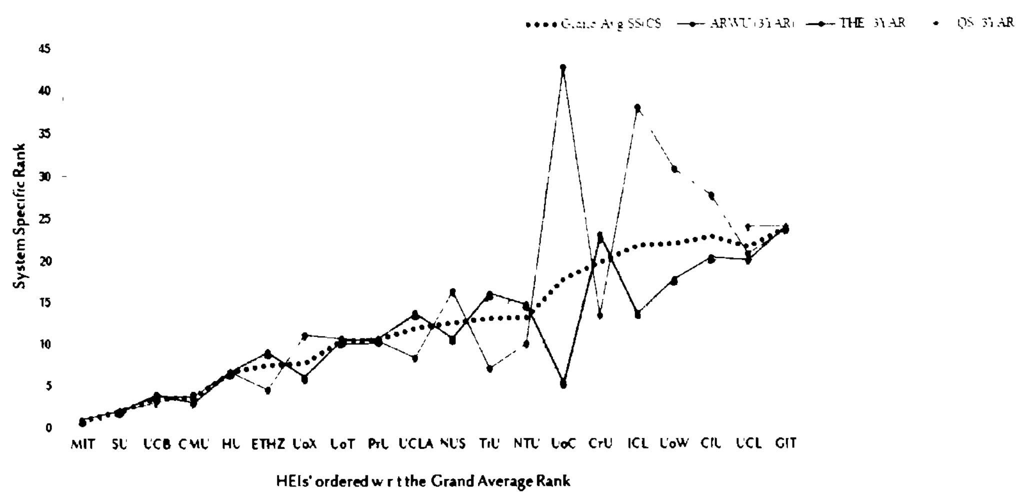


Figure 5.5. Top 20 HEIs w.r.t the Average Ranks by ARWU, THE, QS, and GAR

The result based on the GAR score are presented in Table 4.1 and visualized the statistics in Figure 5.3. According to the graph, the Massachusetts Institute of Technology (MIT) leads the top-performing HEIs with a score of 1.00. Whereas the Georgia Institute of Technology (GIT) has a 23.89 score in the twentieth position. For brevity, a comparison of top-20 HEIs is presented.

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. To get a more reliable academic score for the HEIs, they were analyzed according to the GAR score. In another analysis, we used the GAR score as the benchmark.

In Figure 5.5, the dotted line with black color represents the GAR score, whereas the colored lines represent the average of the ranking scores allotted to the HEIs by each of the three systems — ARWU, THE, and QS in the stipulated three years. It is interesting to note the 3-Year Average Scores allotted by THE and QS are highly similar for the top 20 HEIs. The two systems have a nominal difference with the score allotted by ARWU for the few top-performing HEIs. Nevertheless, the difference becomes significant for HEIs with higher ranking scores. For example, THE and QS allot a similar average score (5.33) to the University of Cambridge, but ARWU allocates a very different average score (42.67) to the HEI.

A similar significant difference exists for Imperial College London (ICL). In general, the difference in the average ranking scores becomes more significant for the HEIs having a rank greater than fifteen. With such differences among the scores by the reputed ranking systems, even in the subject-specific rankings, the end-user would naturally hesitate to believe in an academic ranking system. The differences also provide a clue for criticizing a ranking system in terms of favoritism and commercialization. The analysis highlights the need for a more fine-grained analysis of the academic data and sub-discipline-oriented academic rankings. It is pleasurable observing more details on the web portals of some of the academic ranking systems [23]. Heading toward more transparent and fine-grained rankings is the trend of the ensuing years.

5.4.2. Analysis of the Research Faculty

Agreeing with the opinion that, the single most distinctive competence of a world-class HEI is its ability to attract and retain highly qualified faculty. The analysis explored the strength of influential researchers at the top-ranked HEIs. The results of the analysis are presented in Figure 5.6. In the figure, the bars with green color represent the rank of the HEIs w.r.t the strength of influential research faculty in the Computer Science discipline. 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. The fine-grained analysis shows that even in the top 10 HEIs, some of them lag in having influential faculty in Computer Science. For example, according to the GAR score, Harvard University (HU) is ranked at the fifth position, but it is attaining a ranking score of twenty concerning the computing research faculty.

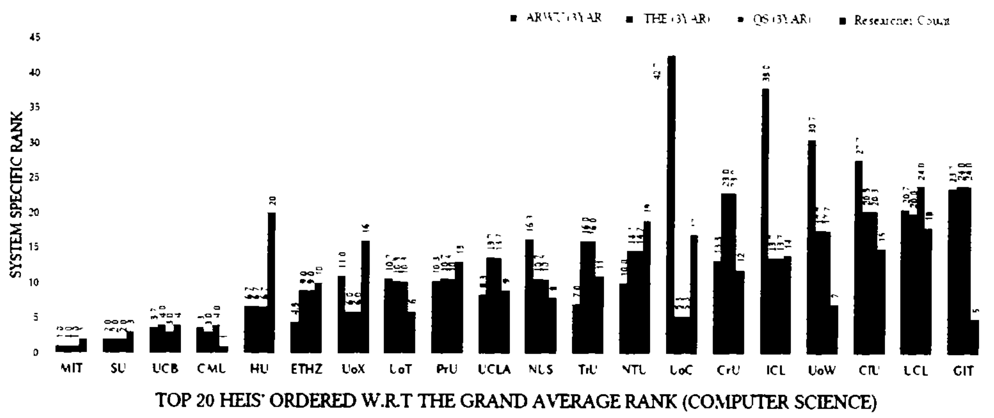


Figure 5.6 Top 20 HEIs ordered w.r.t the GAR and their Research Faculty

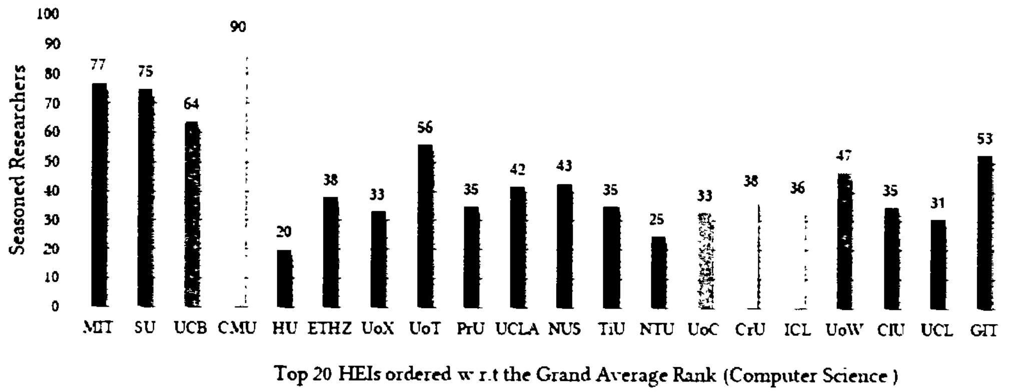


Figure 5.1. Strength of the experienced computing research faculty, in the Top 20 HEIs.

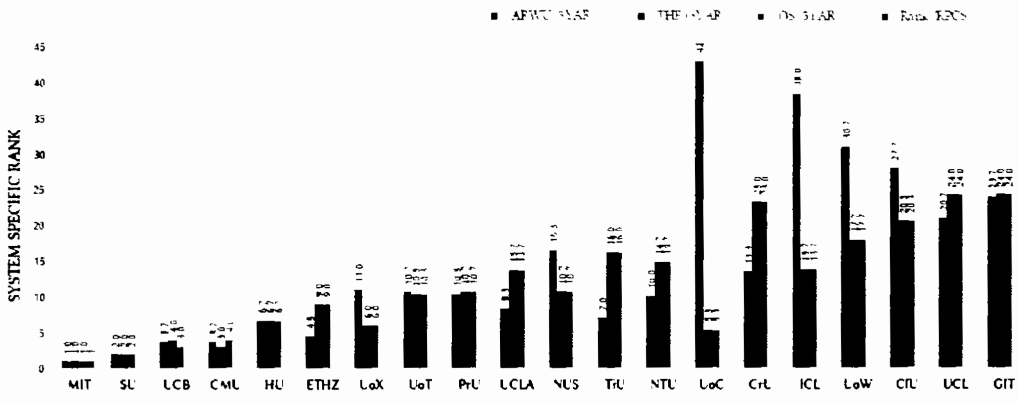


Figure 5.7 Research Productivity (Publication Count) of the Top 20 HEIs

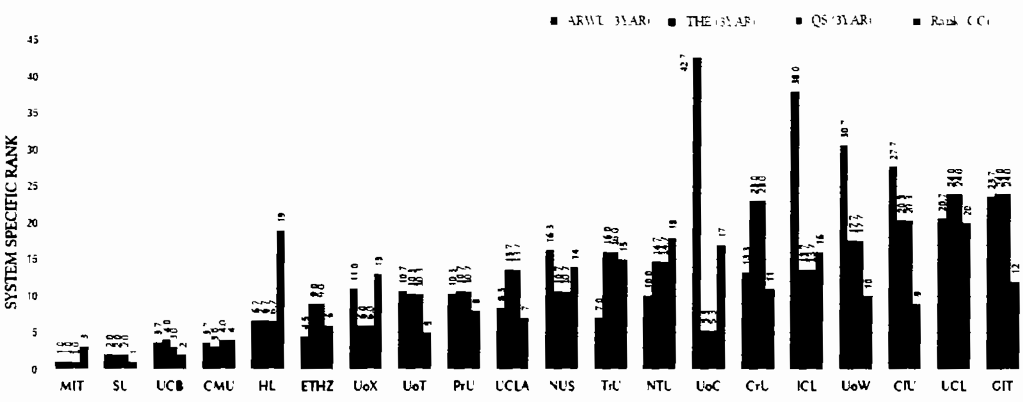


Figure 5.8 Research Impact (Citation Count) of the Top 20 HEIs

Another significant trend exists in cases of the University of Cambridge (UoC), Imperial College London (ICL), and University of Washington (UoW). According to the GAR scores, these institutions are among the top 20 HEIs; Nevertheless, ARWU allocates a considerably lower score (30+) to these institutions. Such differences make it difficult for the audience to believe in their statistics. The need for fine-grained analysis of academic performance becomes more evident. Another way of considering the influential research faculty is gauging the presence of a considerable number of experienced or experienced researchers at an HEI.

The Scopus data repository was explored for this dimension and mapped affiliations of the "World's top 2-percent researchers" in Computer Science, with the top 100 HEIs in the same discipline. A part of the result (for top-20 HEIs) of the analysis is presented in Table 5.3. A researcher was considered "experienced" by having more than 15 years of research experience. For other related statistics, please consult Table 5.3. It is worth mentioning that all the top 20 HEIs (except Harvard University) retained twenty or more experienced researchers belonging to the Computer Science discipline. Carnegie Mellon University (CMU) stood first, having experienced researchers, although it is ranked in fourth position w.r.t the GAR score. Figure 5.7 also highlights the differences between the ranking scores based on GAR and the count of experienced research faculty at other of the top 20 HEIs. The presence of a considerable number of experienced researchers at an HEI is highly attractive for research sponsors and potential research students throughout the world. Nevertheless, this ranking dimension remains hidden under the holistic ranking scores allotted to the HEIs. the analysis suggests a more fine-grained and ranking-dimension-oriented analysis of the HEIs' data.

5.4.3. Analysis of the Research Productivity and Impact

The *Publication Count* (PC) is one of the indicators for representing the *Research Productivity* of a researcher. Nearly all reputed academic ranking systems include the parameter in their ranking methodologies. Therefore, the strength of highly productive researchers at the top-ranked HEIs was explored. The results of the analysis are visualized in Figure 5.7. In the figure, the bars with green color represent the rank of the HEIs w.r.t the *Publication (PC)* of the influential researchers affiliated with the HEI, in the Computer Science discipline. Whereas the other bars represent the 3-year average of the ranking scores allotted to the HEIs, by ARWU, THE, and QS.

The fine-grained analysis shows that even in the top 10 HEIs, some of them lag in having the due research productivity in the Computer Science discipline. For example, according to the GAR score, Harvard University (HU) is ranked at the fifth position, whereas the university is attaining a ranking score of forty-two for the publication count in Computer Science. Other significant differences exist in the cases of the University of Oxford (UoX), National University of Singapore (NUS), University of Cambridge (UoC), Imperial College London (ICL), Nanyang Technological University (NTU), and University College London (UCL). According to the GAR scores, these institutions are among the top 20 HEIs; Nevertheless, ARWU allocates a considerably different score to these institutions. For other related statistics, please consult Table 5.3. The difference exists in their ranks w.r.t the publication count. Such differences make it difficult for the audience to believe in the statistics. The need for fine-grained analysis of academic performance becomes more evident. The constantly flourishing bibliographic data repositories on the Internet, make the analysis possible.

As per academic experts' opinions [69, 2], the prime objective of a higher education institution is the generation of new knowledge benefitting society. That is why gauging the research impact of research work is also among the common objectives of the ranking methodologies of most 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 Scopus data repository, the Citation Count (CC) was extracted. These citations were attained by influential researchers affiliated with the top-performing HEIs in the Computer Science discipline.

A part of the results is visualized in Figure 5.8. For more statistics, please consult Table 5.3. In the figure, the bars with green color represent the rank of an HEIs w.r.t the sum of the *Citation Count* obtained by the influential Computing researchers affiliated with the HEI. 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. As per the graph, the difference between the scores of the Citation Count-based rank and other rankings is nominal for the few top-performing HEIs. Nevertheless, the difference becomes significant for HEIs having positions more than thirteen, w.r.t the GAR ranking. Especially, the scores allocated by ARWU are considerably different for the HEIs. For example, the University of Cambridge (UoC) is unanimously attaining 5.3 as the 3-Year average score by THE and QS; Nevertheless, the same institution is getting a ranking score of 42.7 by ARWU. The HEI is in 17^{th} position w.r.t the Citation Count. Other significant differences exist in the cases of Imperial College London (ICL), and the University of Washington (UoW). A very similar trend can be seen in Figure 5.10, which presents a part of the results of the comparison based on the Citation Count (XSCC) of the researchers, excluding their self-citations. The differences provide a base for the critics of the academic rankings systems.

5.4.4. Considering the Context in the Rankings

The DBpedia data repository maintains various statistics and attributes related to the HEIs. The "Academic Age" was explored for the HEIs with top performance in the Computer Science discipline. The result of the analysis is visualized in Figure 5.10. An overview of the graph reveals the fact that excluding Nanyang Technology University (NTU), all the top 20 HEIs in the Computer Science discipline, with an academic age of more than a hundred years.

It means achieving outstanding status at the world level, which requires decades of consistent effort. Moreover, the HEIs also enjoy conducive settings in terms of their base cities, and countries having stable political and economic conditions. The fine-grained analysis also suggests considering the academic age while comparing the HEIs, even in the subject-specific rankings. 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.

The Scopus data repository maintains the count of research publications by a researcher as a single author or as the first author. These statistics were employed in a parameter – Significant Research Contribution (SRC) of our ResRank methodology. As a part of the intended fine-grained analysis, the SRC was explored for the top-ranked HEIs. A part of the analysis is visualized in Figure 5.11. As per the graph, Carnegie Mellon University (CMU) has more SRC scores than the other top-10 HEIs. Such analysis is also helpful in reducing the difference among the results of reputed ranking systems. For more statistics related to the analysis please consult Table 5.3.

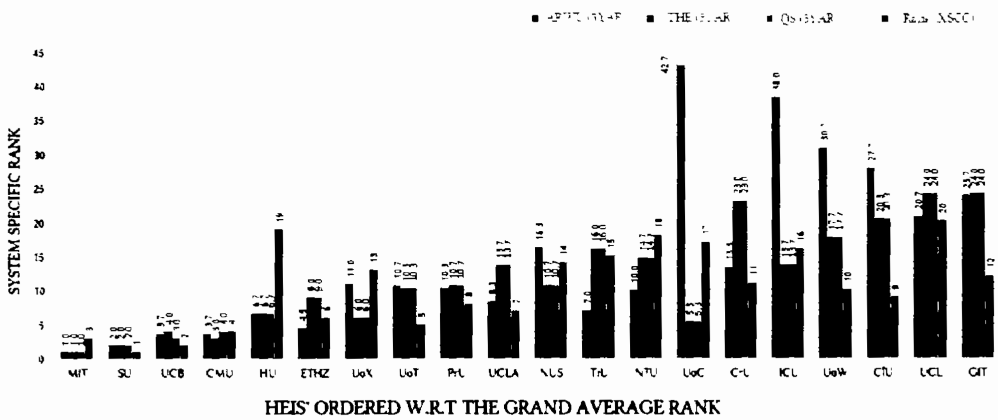


Figure 5.9 Research Impact (Citation Count excluding self-citations) of the Top 20 HEIs

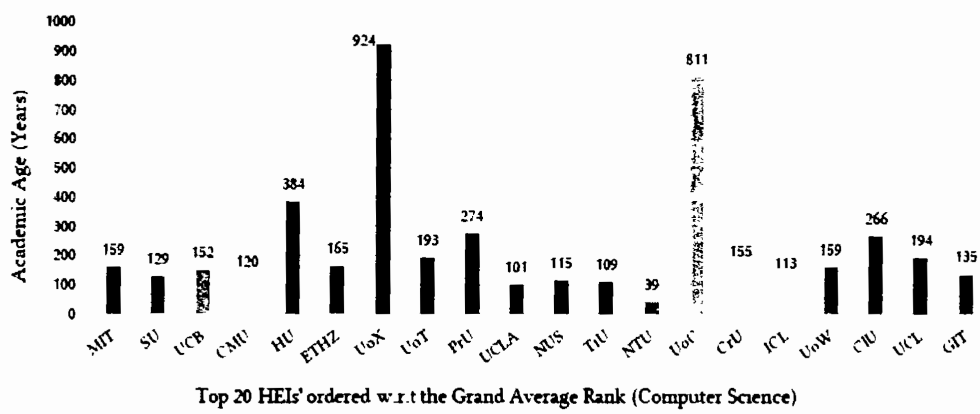


Figure 5.10 Academic Age of the Top 20 HEIs

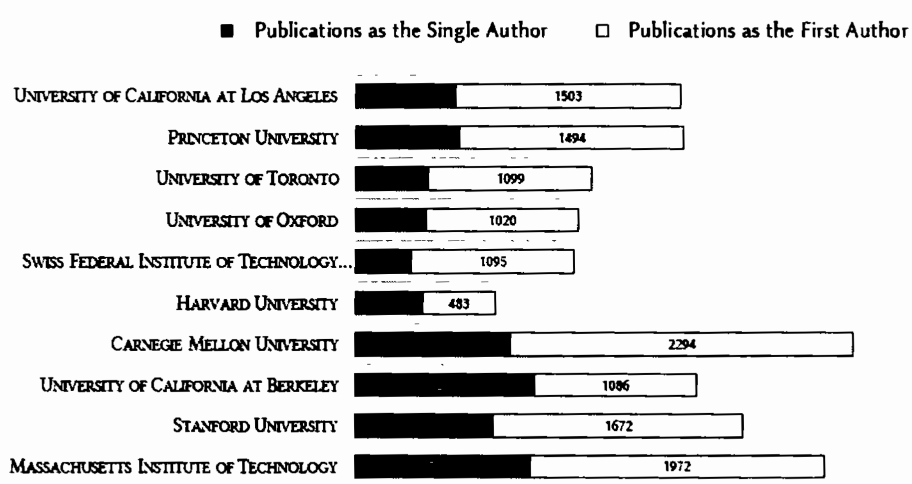


Figure 5.11. Top 10 HEIs w.r.t the faculty with Significant Research Contribution

5.4.5. The Sub-discipline Specific Rankings

Based on the fine-grain analysis the statistics related to the sub-discipline-specific rankings are summarized in table-5 and a part of the results is visualized in Figure 5.12. The figure contains eight graphs highlighting the distribution of the World's Top 2-Percent Researchers (WT2P) in the top 20 HEIs in the computer science discipline. The list of sub-disciplines is based on the categorization adopted in the Scopus data repository.

Among the eight sub-disciplines of computer science, w.r.t the count of researchers, the sub-discipline of ***Artificial Intelligence, and Image Processing*** come first. In this sub-discipline, CMU, MIT, and UoX are at the first three positions respectively, as shown in Figure 5.12 (a). Whereas according to the GAR score the HEIs had 4^{th}, $1^{st,}$ and 7^{th} positions, respectively.

Networking and Telecommunication is the second most significant sub-discipline in terms of count of researchers. In this sub-discipline, MIT, PrU, and (SU along with CrU), are at the first three positions respectively, as shown in Figure 5.12 (b). Whereas according to the GAR score the HEIs had 1^{st}, $9^{th,}$ and (2^{nd}, 15^{th}) positions, respectively. The difference in WT2P researchers' count is five between the first two HEIs, however, the average change is one, among all other HEIs.

Software Engineering is the third most influential sub-discipline w.r.t the count of researchers. In this sub-discipline, UoW, CMU, and GIT are at the first three positions respectively, as shown in Figure 5.12 (c). Nevertheless, according to GAR score the HEIs had 16^{th}, $3^{rd,}$ and 20^{th} positions, respectively. For it, the average change in the WT2P researcher count is not significant.

Computer Hardware and Architecture is the next sub-discipline in the queue. Here, CMU, SU and UoT clinched the first three positions, as shown in Figure 5.12 (d), however, according to GAR score the institutions had 4^{th}, $2^{nd,}$ and 8^{th} positions, respectively. With the negligible change in average WT2P researchers' count among other HEIs, the first four HEIs retained most of the influential researchers.

Computer Theory and Mathematics come next in the disciplines. In it, MIT, PrU, and (SU along with CrU), are at the first, second, and third positions respectively, as shown in Figure 5.12 (e). Whereas according to the GAR score these instructions had 1^{st}, 2^{nd}, and (9th, 15th) positions, respectively. In this discipline, the average variation in the WT2P researchers' count has significant variation. Moreover, some of the top 20 HEIs have no influential researchers in this discipline.

Information Systems come next in the disciplines. In this sub-discipline, according to Figure 5.12 (f), MIT along with SU achieve the first position, UCB, UoT, and NUS collectively obtained the second position, whereas four HEIs were in the third position. According to the GAR score, these instructions have very different ranks in the Computer Science discipline than their ranks in the sub-discipline.

In the ***Medical Informatics*** sub-discipline, as shown in Figure 5.12 (g), the ranks of the top-20 HEIs w.r.t WT2P researchers' count, are amazing. The two HEIs (UoW and CIU) having the first position, are in the 17th and 18th positions respectively according to the GAR score. Moreover, 16 out of the top 20 HEIs, have no WT2P researcher in the sub-discipline!

Distributed Computing comes in the last of the eight sub-disciplines. In it, GIT stood first, nevertheless according to the GAR score, it is in 20th position! Moreover, 13 out of the top 20 HEIs, have no WT2P researcher in the sub-discipline.

A product of the research analysis is the sub-discipline-specific ranking of institutions, according to their ability to hire and retain WT2P researchers in the computer science discipline. A part of the results is presented in Table 5.4, The table contains a list of the top-five institution in each of the above-mentioned eight sub-disciplines. Quite expectedly, the lists made visible many new institutions, that are usually not visible, among the top 20 HEIs, in ranking tables of the reputed academic ranking systems. It is worth mentioning, along with the degree awarding HEIs, many intuitions, companies, labs, and independent research centers become visible, as shown in Table 5.4, in terms of research faculty, impact, and contribution. The companies like Microsoft Research and Google LLC are the leaders in AIIP and SE.

Using the *ResRank* methodology employed in the research work, the affiliations of the top 5000 researchers were grouped to find the leading institutions having a significant number of computing researchers. While doing the analysis, the GAR score was not considered. A part of the results of the analysis is presented in appendix (b), which lists the top 50 institutions ranked according to the *ResRank* methodology.

Table 5.4. Top 5 research institutions, in the eight sub-disciplines

Sub Discipline	Institution	Researchers' Count
Artificial Intelligence & Image Processing	**Microsoft Research**	52
	Google LLC	50
	Carnegie Mellon University	49
	University of Oxford	27
	University of California, Berkeley	26
Computation Theory & Mathematics	**Tel Aviv University**	9
	Massachusetts Institute of Technology	8
	Princeton University	8
	Cornell University	7
	Rutgers University-New Brunswick	7
Computer Hardware & Architecture	**The University of Illinois at Urbana-Champaign**	11
	Carnegie Mellon University	10
	University of Michigan, Ann Arbor	9
	The University of Texas at Austin	8
	Purdue University	6
Distributed Computing	Georgia Institute of Technology	7
	The University of Illinois at Urbana-Champaign	7
	IBM Thomas J. Watson Research Center	6
	Lawrence Berkeley National Laboratory	6
	Argonne National Laboratory	4
Information Systems	**Microsoft Research**	9
	Hong Kong University of Science and Technology	5
	Stanford University	5
	The University of Arizona	5
	City University of Hong Kong	4
Medical Informatics	**Maastricht University**	6
	National Library of Medicine	5
	McGill University	4
	Oregon Health & Science University	4
	Vanderbilt University Medical Center	4
Networking & Telecomm	Stanford University	29
	Nokia Bell Labs	27
	The University of Illinois at Urbana-Champaign	25
	The Ohio State University	24
	University of Michigan, Ann Arbor	24
Software Engineering	**Microsoft Research**	13
	Google LLC	9
	University of Washington, Seattle	9
	Carnegie Mellon University	8
	Georgia Institute of Technology	7

A name in the bold font indicates that the institution is among the top 20 (w.r.t the ResRank Methodology) but it is not among the top 20 w.r.t the Grand Average Rank (GAR).

According to the list, eight new research institutions secured their positions among the top 20 leading institutions. It is noteworthy that they were not part of the list containing the top 20 institutions according to rankings of ARWU, QS, and THE. The list in the appendix shows that, besides HEIs, certain research companies and laboratories are having an outstanding number of influential researchers in the Computer Science discipline. Moreover, six new HEIs and two research companies are added to the list.

The sub-discipline-specific analysis reveals the fact that the subject-specific (discipline-specific) rankings published by the reputed academic rankings, do not present the insight at the sub-discipline level, despite due importance of the granularity. Alternatively, the sub-discipline-specific rankings demonstrate that various HEIs not visible on the top, in the lists of reputed academic ranking systems, perform remarkably well in the sub-disciplines.

The evolution in the academic ranking process is leading toward sub-discipline-specific rankings. The academic world is offering examples. Institutions, like Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates[20] one example. Soon the potential users of academic rankings, including HEI's policymakers and managers, along with the prospective students shall demand sub-discipline-specific rankings. The growing interest in subject-specific and sub-discipline-specific rankings is irreversible and inevitable. The reputed International rankings would be the first to expand the limited picture of global higher education.

The DBpedia data repository provided various meaningful information about the countries and HEIs including the base country of an HEI, its exact location, city, year of establishment, etc. A part of the result is presented in Figure 5.13. In the figure, the green bars represent the count of researchers belonging to the Computer Science discipline, while the yellow bars represent the value of the Grand Domestic Product (GPD in 100 million USD) of the top 15 countries.

[20] mbzuai.ac.ae [Accessed on Nov 2021]

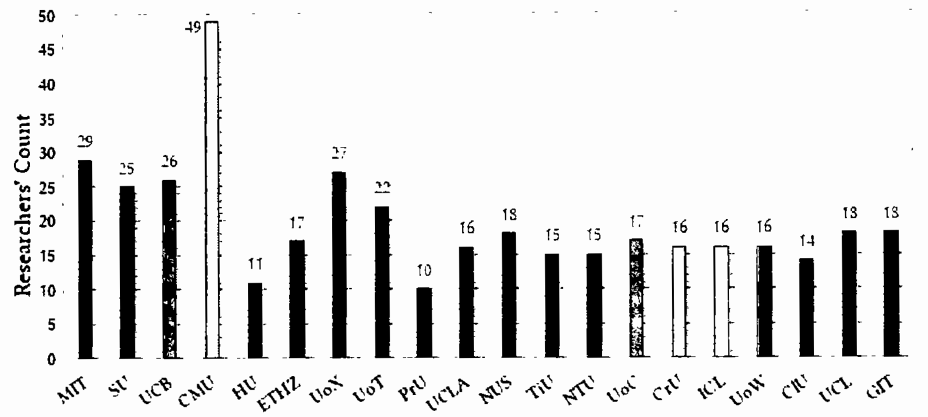
Using that, the countries were identified that have a considerable number of the World's Top-2 Percent Researchers in the discipline of Computer Science. Quite understandably, the top ten countries are the most stable and prosperous courtiers. Governments of developed countries are cognizant of the role of HEIs and talented academicians, in the advancement of their socioeconomic development. Thus, they nurture research, innovation, and commercialization through the institutions. The data visualized in Figure 5.13 also contains the percentage of the Human Development Index (HDI) score, on the secondary axis. The trend line emphasizes the fact that being exceptional in terms of human prosperity does not guarantee excellence in research and innovation. However, the USA and UK are demonstrating the need for human prosperity for better performance. Moreover, the GPDs of the top 15 countries, is the leading economic powers in the World. These countries are prosperous due to their knowledge-based economies catalyzed by research and innovation in science and technology.

It is also worth mentioning that comparisons of HEIs having very different sizes, budgets, locations, etc. also raise the question of the validity of the ranking process. It is good to observe that leading academic ranking systems are proceeding with 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 the such fine-grained analysis. The companies like Microsoft Research and Google LLC are the leaders in AIIP and SE. It also reflects the attention of authorities to consider such entities in the ranking systems as well.

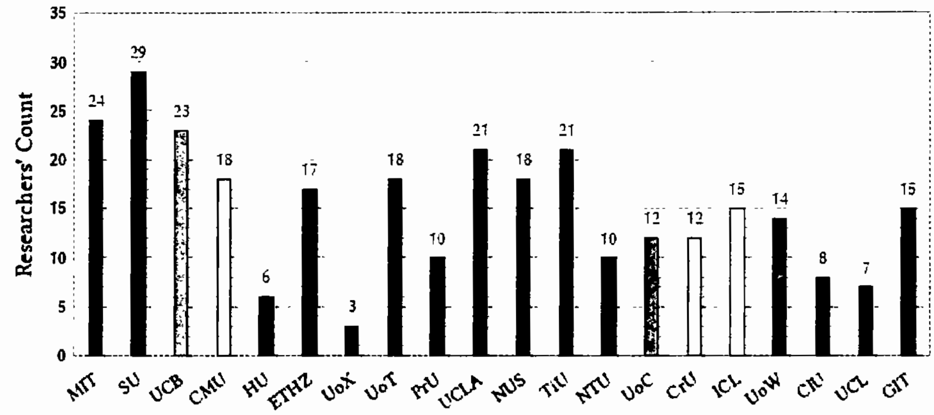
5.5. Conclusions

Global HEI rankings face various criticisms, but they are here to stay. Reputed ranking entities have improved the quality of their rankings substantially during the past fifteen years. However, based on our analysis, it can be argued that they must address the growing interest in the fine-grained assessment of data related to academic rankings. Our analysis revealed some of the factors causing substantial variances between the rankings scores of HEIs published by various reputed rankings entities. The holistic ranking scores hide many valuable achievements of an HEI, even in the subject-specific rankings.

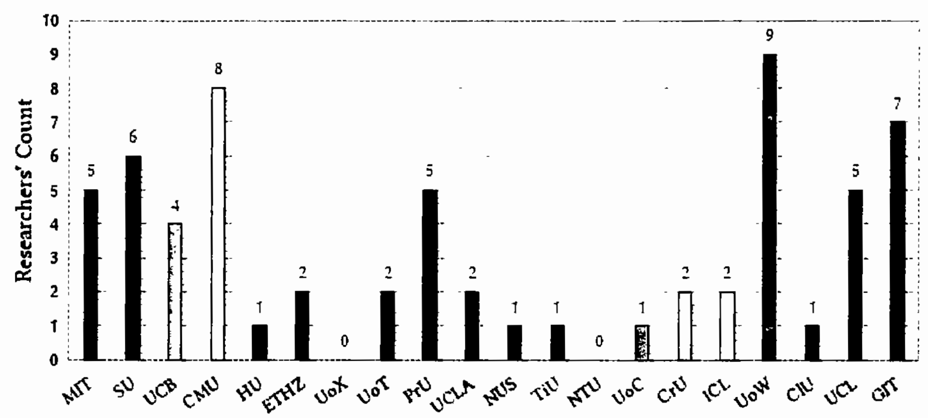
a) Artificial Intelligence and Image Processing



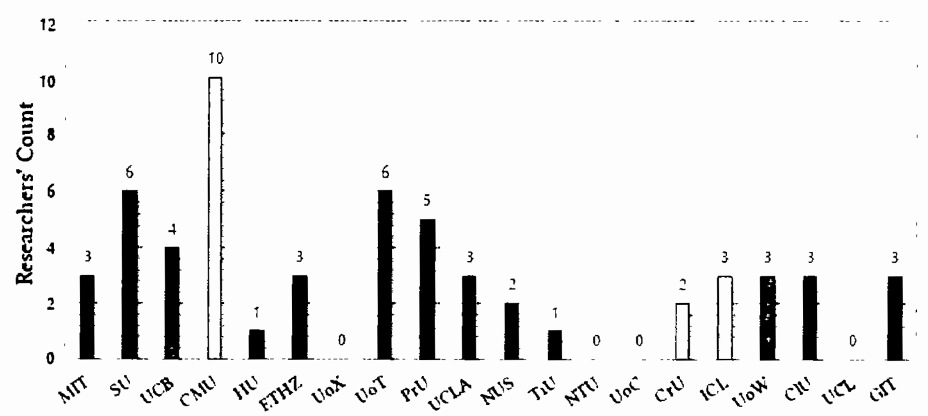
b) Networking and Telecommunication



c) Software Engineering



d) Computer Hardware and Architecture



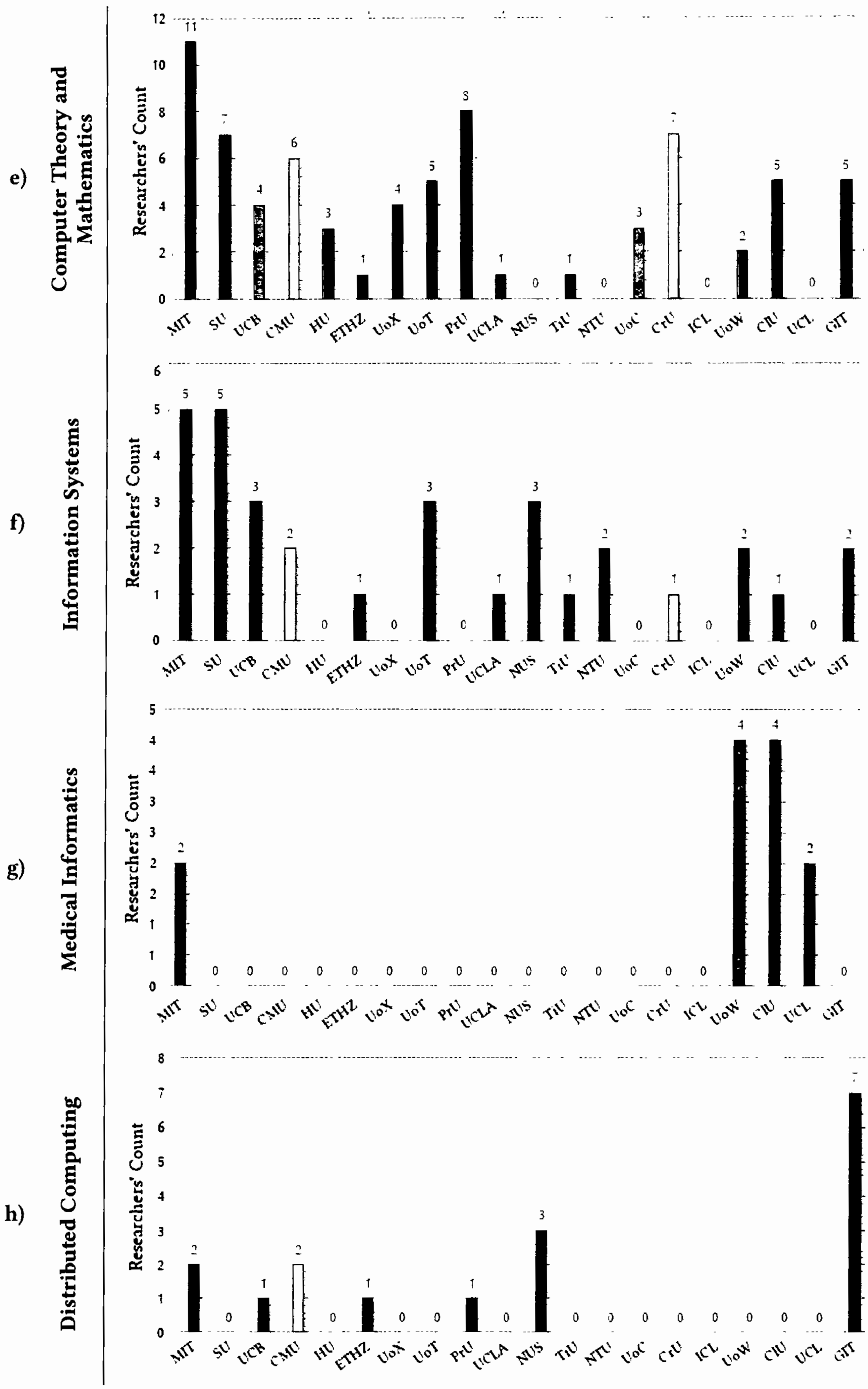


Figure 5.12 (a to h). The rank of Top 20 HEIs in the Sub-disciplines of Computer Science

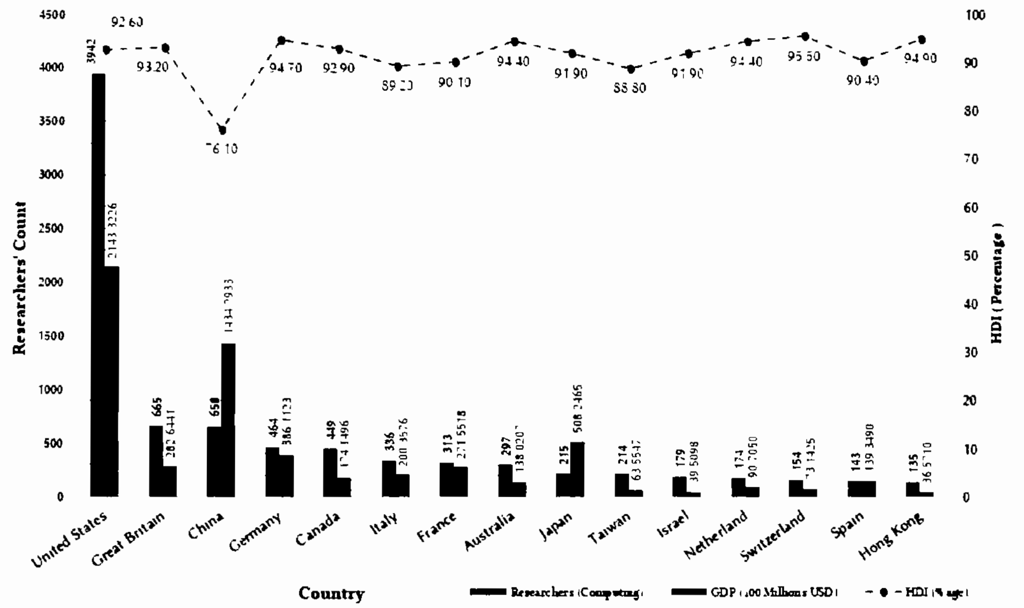


Figure 5.13. Top 15 Countries w.r.t the World Top-2 Percent Researchers

The sub-discipline-specific ranking based on the fine-grained analysis highlighted that various HEIs perform remarkably well at the sub-discipline level, although they do not feature on the top of the lists produced by well-known academic ranking entities. Sub-discipline level will lead to fairer and better insight in ranking. The *ResRank* ranking methodology was applied, instead of the PageRank algorithm or OpenRank ranking methodology mentioned in the related work to make the analysis more fine-grained. The reputed international ranking entities have the potential to make their ranking methodologies more fine-grained and transparent. Considering the context of an HEI is also important while including it in the ranking process. The contextual factors, including the economic condition of the base country, academic age, languages, and size, have a significant effect on the position of an HEI in the ranking lists. Overlooking this context would result in biased judgments.

5.6. Summary

After attaining maturity in the holistic ranking scores, the reputed ranking entities of the global HEIs have focused on the subject-specific and regional rankings. Although rankings face various critics, they are evolving and undoubtedly here to stay. This research work is a part of the fine-grained analysis of the academic rankings focusing on the computer science discipline. Earlier, two related research efforts has been done and their results were reported in reputed research journals. For more detail, please consult chapters three and four. In this research work, the fine-grained analysis employs the highly curated data produced and published by the three well-known international academic rankings – ARWU, THE, and QS in the years 2018, 2019, and 2020. The researchers' profiles are obtained from the Scopus repository, while the DBpedia repository is employed for the information about HEIs and their base countries.

The subject-specific rankings in computing, are analyzed according to the *research productivity*, *research impact,* and *research contribution* of the researchers affiliated with the top-ranked HEIs. In the analysis, the Grand Average Rank (GAR) score and the *ResRank* measure were employed for understanding the causes of the differences among the results of the ranking entities and to produce the sub-discipline-specific academic rankings. The fine-grained ranking results provided better insight into the academic performance of the global HEIs in various sub-disciplines of Computing.

An innovative and interesting part of this research effort is the generation of the *sub-discipline-specific* academic ranking. The fine-grain analysis of the existing academic rankings reveals the fact that the subject-specific rankings hide the insight at the sub-discipline level. At the sub-discipline level, various HEIs perform remarkably well, although they are not visible at the top, in the lists of reputed academic ranking systems. Such insights would be more decision-supportive for the end-user due to its expanded picture of the global HEIs.

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.

Chapter 6

Conclusions and Future Directions

This chapter summarizes the whole research effort, with a brief description of the research domain, its major research issues, the focused research problem, and the proposed solution. A brief description of the fine-grained analysis approach, data sources, research findings, and research implications, is also a part of it.

6.1. The Research Effort – An Overview

Educational institutions, especially those dealing with higher education, are considered catalysts of the national and global development process. As, they enhance a nation's competitive advantage and stimulate the development process by nurturing research, innovation, and commercialization. The rankings have gained substantial acceptance by major stakeholders – academicians, prospective students, and employers. The ranking process has considerably evolved in the past fifteen years. Starting with the holistic rankings of world universities in 2003, it has crossed the milestone of subject-specific rankings. Currently, various ranking systems are gauging and publishing the HEI's rankings, including SRC, QS, and THE. These systems significantly have improved their quality during the past fifteen years. Nevertheless, academic ranking experts have pointed out multiple debatable issues in the academic 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

This research work focused on the enhancement of the creditability of the ranking process, by employing quantitative ranking indicators based on publicly verifiable data sources and fine-grained analysis of the target data. To address the focused research problem, a fine-grained analysis of the objective ranking indicators extracted from public data sources was performed. The analysis was done at the level of ranking indicators i.e., the *research productivity, research impact, and research contribution of the influential researchers* affiliated with the top-ranked HEIs. Two new rankings methodologies – *OpenRank* and, *ResRank* were designed to achieve the desired ranking results.

The *OpenRank* methodology ranked the global HEIs using the objective indicators extracted from two well-known publicly verifiable data repositories — the ArnetMiner and DBpedia. The results of the methodology demonstrated the feasibility of more transparent and reliable academic rankings. The *ResRank* methodology was used to identify influential researchers in various scientific disciplines. While, for consistent comparison of the subject-specific rankings of global HEIs, the *Grand Average Rank* (GAR) metric is employed.

Following the fine-grained analysis approach, the data was explored at the ranking indicator and dimension levels including the *research productivity, research impact,* and *research contribution* of the influential research scholars. The analysis employed highly curated data published by the three well-known international academic rankings – ARWU, THE, and QS in the years 2017—2020. The researchers' profiles were obtained from the Scopus, Google Scholar Citations, and ArnetMiner repositories. The DBpedia repository was used to retrieve information about HEIs and their locations.

The sub-discipline-specific academic rankings were also an outcome of the fine-grained analysis. These rankings provided a better insight into the performance of an HEI; thereby becoming more decision supportive. The sub-discipline-specific academic ranking in the Computing domain is among the first few such efforts.

6.2. Research Findings

The intended research work was completed in three phases and the results of each phase were reported in various research journals. A comprehensive description of each of the phases is given in chapters 3, 4, and 5. Research findings of each phase are also described at the end of the chapters. Nevertheless, for a quick overview, the findings are summarized here.

a) More transparent and reliable ranking results are achievable

Global HEIs rankings face various critics, but they are here to stay. The use of objective and publicly verifiable rankings indicators makes the rankings results more transparent, representative, and reliable. Whereas, giving substantial weight to subjective indicators leave more space for favoritism. The proposed Open-Rank methodology emphasizes the employment of objective indicators, that can be verified scientifically and are publicly available. It also gives substantial weightage to both research productivity and the academic contribution of an HEI. This comprehensive approach broadens the coverage of HEIs, especially the newly established HEIs. The resultant rankings were evaluated and equated with rankings produced by the two reputed academic ranking systems, ARWU rankings, and QS University rankings in 2017. Evaluation of the methodology advocates its effectiveness and quick reproducibility with the low cost of data collection.

The use of diverse data sources facilitates a comprehensive understanding of the factors that catalyze higher education. The resultant academic rankings make the ranking process more transparent and more representative, thereby the ranking results less controversial. The data sources enabled the OpenRank methodology to explore new ranking indicators like the count of publications of an HEI on the favorite venues (top 50 venues recorded in the dataset, having more than 3500 publications). The DBpedia contains information about the geographical regions and countries of the world. The methodology also explored the relationship between the Human Development Index (HDI) of a country and the number of HEIs owned by it. Interestingly, the trend highlights the fact that outstanding human prosperity does not ensure an outstanding share in research and innovation – nevertheless, the United States of America and the United Kingdom advocated prosperity for better performance. The experiments also explored the relationship between the Grand Domestic Product (GDP) of a country and the number of HEIs owned by it. The United States of America and the United Kingdom again outperformed in hosting the top-performing HEIs, as compared to various other countries with a higher GDP.

It is worth mentioning that although, the academic data available on the Internet is abundant and perpetually increasing. Nevertheless, most of the data repositories are not purposely built for academic ranking. DBpedia and Arnet Miner repositories also have the problem. So, data quality remained a problem. In the future, more reliable and optimal results would be produced by employing more enriched and curated data sources.

b) The fine-grained analysis makes the ranking results more representative

The academic ranking process has considerably evolved in the past fifteen years; Currently, they are more informative and user-friendly than they were a decade ago. Nevertheless, the academic rankings published by the reputed ranking entities are facing various criticism, in terms of their transparency, validity, and coverage. Similarly, 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. The analysis according to various ranking dimensions (such as Academic Quality, Research Productivity, Graduate-Employability, Industry-Academia Linkage, and Internationalization.) reveals the factors, causing significant differences among the rankings scores of HEIs. Exploration of various ranking dimensions can offer a better understanding of educational dynamics. The fine-grained analysis would better answer the needs of prospective students and academic management.

c) The sub-discipline-specific academic rankings are needed

The sub-discipline-specific ranking based on the fine-grained analysis highlighted the fact that various HEIs perform remarkably well at the sub-discipline level, although they are not present on the top of the lists produced by well-known academic ranking entities. The reputed international ranking entities have the potential to make the ranking methodologies more fine-grained and transparent. The sub-discipline-specific academic rankings would make the rankings score more representative, and the ranking process less controversial. The analysis also ranked HEIs in the following eight sub-disciplines of Computer Sciences.

- Artificial Intelligence and Image Processing
- Networking and Telecommunication
- Software Engineering
- Computer Hardware and Architecture
- Computer Theory and Mathematics
- Information Systems
- Medical Informatics
- Distributed Computing

d) Context-aware academic rankings would be more representative

Considering the context of an HEI is also important while including it in the ranking process. The contextual factors including the economic condition of the base county, academic age, languages, and size significantly affect the position of an HEI in the ranking lists. Overlooking the context would result in biased judgments. The reputed international ranking entities have the potential to make their ranking methodologies more fine-grained and transparent.

6.3. Research Implications (Future Directions)

The findings of the research work highlight various implications. A brief description of the implications is given below.

a) Improvement in the accuracy of the fine-grained analysis

In the future, improvement in the accuracy of the fine-grained analysis would be focused on employing more relevant public data sources including the bibliographic data maintained by Association for Computing Machinery (ACM) and Scopus, Microsoft Academics, 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 should be analyzed in the ranking dimensions. The reputed academic ranking entities would be the first to expand the limited picture of global higher education.

b) Sub-discipline-specific rankings of other scientific domains

Following the pattern adopted in this research work, *Sub-discipline-specific of* other scientific domains such as Engineering, Economics, Business Administration, and Physical Sciences would be produced.

c) Exploration of various ranking dimensions

There are other ranking dimensions like the *affiliation of influential research faculty* with an HEI, and *Graduate Employability,* that can be explored for a better understanding of the educational dynamics. Their fine-grained analysis would better serve the needs of prospective students and academic management.

d) Research recommender system for the prospective students

A research recommender system for novice researchers and prospective students can also be developed by employing objective and publicly available data sources. We are also working on a Subject-Specific Ranking of highly influential researchers and global HEIs based on publicly verifiable and objective data sources including Google Scholar's Citations and Microsoft Academic Graph.

e) Need for a purposefully built, verifiable electronic data source

To enable better performance evaluation of global HEIs, the academic world needs a purposefully built, verifiable electronic data source. The creation of such a useful data repository would enable the stakeholders to improve academic planning, monitoring, and assessment. More reliable, transparent, and less controversial academic rankings would be produced using the data source.

References

[1] S. Friso, N. Cameron, H. Chun-Kai and G. Paul, "A Longitudinal Analysis of University Rankings," *Quantitative Science Studies*, vol. 1, no. 3, pp. 1109-1135, 2020.

[2] I M. Lokman, "Highly prestigious international academic awards and their impact on university rankings," *Quantitative Science Studies*, vol. 1, no. 2, pp. 824-848, 2020.

[3] M.-H. Huang, "A comparison of three major academic rankings for world universities: From a research evaluation perspective.," *Journal of Library & Information Studies*, vol. 1, no. 9, pp. 1-25, 2011.

[4] B. Millot, "International Rankings: Universities vs. Higher Education Systems," *International Journal of Educational Development*, vol. 40, pp. 156-165, 2015 http://dx.doi.org/10.1016/j.ijedudev.2014.10.004.

[5] R Garcia and F. Herrera, "An Insight into the Importance of National University Rankings in an International Context," in *ISSI Conference*, Vienna, 2013.

[6] A Rad, B. Naderi, and M. Soltani, "Clustering and Ranking University Majors Using Data Mining and AHP Algorithms. A Case Study in Iran," *Expert Systems with Applications*, vol. 38, p. 755--763., 2011.

[7] I. F. Aguillo and J. Bar-Ilan, "Comparing University Rankings," *Journal of Scientometrics*, vol. 85, pp. 243-256, 23 Feb 2010

[8] G. B. M. Olcay, "Is Measuring The Knowledge Creation of Universities Possible?: A Review of University Rankings," *Technol. Forecast. Soc. Change*, 2016. http://dx.doi.org/10 1016/j.techfore.2016.03.029.

[9] L. d. Mesnard., "On Some Flaws of University Rankings: The Example of the Scimago Report," *Journal of Socio-Economics*, vol. 41, p. 495– 499, 2012.

[10] G. Boulton, "University Rankings: Diversity, Excellence and the European Initiative," *Science Direct. Procedia Social and Behavioral Sciences*, vol. 13, p. 74–82, 2011.

[11] N C. Liu, "ARWU-Academic Ranking Methodology," 2014. [Online]. Available: http://www.shanghairanking.com/ARWU-Methodology-2014.html. [Accessed 03 2016].

[12] Q. S. Writer, "QS World University Rankings Methodology," 11 Sep 2015. [Online]. Available: http://www.topuniversities.com/university-rankings-articles/world-university-rankings/qs-world-university-rankings-methodology?page=1. [Accessed 25 Mar 2016].

[13] T. O. Writer, "THE Ranking Methodology 2016," 2016. [Online]. Available: https://www.timeshighereducation.com/news/ranking-methodology-2016. [Accessed 10 Aug 2016].

[14] C.-S. Lin, M.-H. Huang and D.-Z. Chen., "The Influences of Counting Methods on University Rankings Based on Paper Count and Citation Count," *Journal of Informetrics*, vol. 7, p. 611–621, 2013. http://dx.doi org/10 1016/j.joi.2013.03.007.

[15] R. Meymandpour and J G. Davis, "Ranking Universities Using Linked Open Data," in *International Workshop on the Linked Data on the Web (LDOW'13)*, 2013.

[16] J. Tang, J Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "ArnetMiner: Extraction and Mining of Academic Social Networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, USA, 2008.

[17] ArnetMiner, "Citation Network Dataset," [Online]. Available: https://www.aminer.cn/citation. [Accessed December 2017].

[18] ArnetMiner, "Open Academic Graph," [Online]. Available: https://www.aminer.cn/open-academic-graph. [Accessed 14 November 2019].

[19] ArnetMiner, "Open Data and Codes by Arnetminer," [Online]. Available: https://www.aminer cn/download. [Accessed 12 November 2019].

[20] ArnetMiner, "Organization Ranking," [Online]. Available: https://www.aminer.org/ranks/org. [Accessed 15 September 2019]

[21] C.-K. Huang, C. Neylon, C. Brookes-Kenworthy, R. Hosking, L. Montgomery, K. Wilson, and A. Ozaygen, "Comparison of bibliographic data sources: Implications for the robustness of university rankings," *QSS*, vol. 1, no. 2, pp. 445-478, January 2020.

[22] J. Baas, M. Schotten, A. Plume, G. Côté and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies.," *Quantitative Science Studies*, vol. 1, p 377–386, 2020.

[23] A. Martín-Martín, M. Thelwall, E. OrdunaMalea and E. Delgado López-Cózar, "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations," 2018.

[24] J. P. A. Ioannidis, J Baas, R. Klavans and K. W. Boyack, "A standardized citation metrics author database annotated for scientific field," *PLOS BIOLOGY*, vol. 17, no. 8, pp. 21-27, August 2019.

[25] Isidro F AGUILLO, *Ranking Web or Webometrics - Academic Ranking Methodology*, 2014.

[26] N. Kapur, N. Lytkin, B.-C. Chen, D. Agarwal and I. Perisic, "Ranking universities based on career outcomes of graduates," in *KDD '16, August 13 - 17, 2016*, San Francisco, CA, USA, 2016.

[27] F. Koen, J H. Gaston, and H. Jarno, "What drives university research performance? An analysis using CWTS Leiden Ranking data," *Journal of Informetrics*, vol. 11, no. 3, pp. 859-872, 11 2017.

[28] P. Haddawy, S. U Hassan, C. W. Abbey, and L. I. Beng, "Uncovering fine-grained research excellence· The Global research Benchmarking System," *Journal of Informetrics*, vol. 11, p. 389–406, 2017.

[29] Y. Dong, H. Ma, Z. Shen, and K. Wang, "A Century of Science: Globalization of Scientific Collaborations, Citations, and Innovations," in *KDD '17, August 13-17, 2017*, Halifax, NS, Canada, 2017.

[30] J. L. Ortega, "Influence of co-authorship networks in the research impact Ego network analyses from Microsoft Academic Search," *Journal of Informetrics*, vol. 8, p. 728–737, 2014.

[31] A. Sinha, S. Zhihong, S. Yang and M. Hao, "An Overview of Microsoft Academic Service (MAS) and Applications," in *ACM 978-1-4503-3473-0/15/05.*, Florence, Italy., 2015.

[32] S. N.K., S. Mathew K., and S. Cherukodan, "Impact of scholarly output on university ranking," *Global Knowledge, Memory and Communication*, vol. 67, no. 3, pp. 154-165, 2018.

[33] B. Yang, R. Ronald, H. Shuiqing and Y. Sulan, "Do first rate scientists work at first rate organizations?," *Malaysian Journal of Library & Information Science*, vol. 20, no. 1, p. 47–60., 2015.

[34] A. Giovanni and A. D. Ciriaco, "Ranking research institutions by the number of highly-cited articles per scientist," *Journal of Informetrics*, vol. 9, pp. 915-923, 2015.

[35] M. S. Qureshi, A. Daud, M. K. Hayat, and M. T. Afzal, "OpenRank - A Novel Approach to Rank Universities Using Objective and Publicly Verifiable Data Sources," *Library Hi-Tech*, vol. 39, no. 1, 2021.

[36] J Mingers, J. R. O'Hanley, and M. Okunola, "Using Google Scholar institutional level data to evaluate the quality of university research," *Scientometrics*, vol. 113, p 1627–1643, 2017.

[37] N. P. Fredrik and S. Gunnar, "How can differences in international university rankings be explained? " *Scientometrics*, vol. 109, no. 3, p. 2263–2278, 2016.

[38] Z. Jun, X. Bo, L. Jiaying, T. Amr and Z. Al, "PePSI: Personalized Prediction of Scholars' Impact in Heterogeneous Temporal Academic Networks," *IEEE ACCESS*, 2018.

[39] B. Uslu, "A path for ranking success: what does the expanded indicator-set of international university rankings suggest?," *Higher Education*, vol. 80, p. 949–972, 2020.

[40] M. McAleer, T. Nakamura and C. Watkins, "Size, Internationalization, and University Rankings: Evaluating and Predicting Times Higher Education (THE) Data for Japan," *Sustainability*, vol. 11, 2019

[41] M. S Qureshi and A Daud, "Fine-grained academic rankings: mapping affiliation of the influential researchers with the top ranked HEIs," *Scientometrics*, vol. 126, p. 8331–8361, September 2021.

[42] F A. Massucci and D Docampo, "Measuring the academic reputation through citation networks via PageRank," *Journal of Informetrics*, vol. 13, p. 185–201, 2019.

[43] J Ioanidis, K. W Boyack and J. Baas, "Updated science-wide author databases of standardized citation indicators," *PLoS Biol*, vol. 18, no. 10, pp. 15-17, October 2020.

[44] J. Ioannidis, J Baas, R. Klavans and K. W. Boyack, "A standardized citation metrics author database annotated for scientific field," *PLoS Biol*, vol. 17, no. 8, August 2019.

[45] A. Giovanni , A. D. Ciriaco and S. Anastasiia, "The ratio of top scientists to the academic staff as an indicator of the competitive strength of universities," *JournalofInformetrics*, pp. 596-605, 2016.

[46] G. Abramo, C. Andrea D'Angelo, and F. Di Costa, "The collaboration behavior of top scientists," *Scientometrics,* vol. 118, p. 215–232, November 2018.

[47] C. G. McCoy, M. L. Nelson, and M. C. Weigle, "Mining the Web to approximate university rankings," *Information Discovery and Delivery,* vol. 46, no. 3, p. 173–183, 2018.

[48] M. A. Francesco and D. Docampo, "Measuring the academic reputation through citation networks," *Journal of Informetrics,* vol. 13, p. 185–201, January 2019.

[49] S Sternberg, *Celebrating 80 Years - A Timeline of Events in The Life of U S. News & World Report,* 2013.

[50] Federkeil, "Rankings and Quality Assurance in Higher Education," *Higher Education in Europe,* vol. 33, no. 2, pp. 219-231, 2011. URL: http://dx.doi.org/10.1080/03797720802254023.

[51] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top-k Lists.," in *Fourteenth annual ACM-SIAM symposium on Discrete algorithms.,* Baltimore, Maryland, 2003.

[52] W. Webber, A. Maffat and J. Zobel, "A similarity measure for indefinite rankings.," *ACM Transactions on Information Systems,* pp. 1-38, 28 4 2010.

[53] G. Abramo, D. C Andrea, and A. Soldatenkova, "The ratio of top scientists to the academic staff as an indicator of the competitive strength of universities," *JournalofInformetrics,* vol. 10, p. 596–605, 2016.

[54] A. O. Jo, "The "Dark Side" of Academics? Emerging Issues in the Gaming and Manipulation of Metrics in Higher Education," *The Review of Higher Education,* vol. 42, no. 3, pp. 859-877, 2019.

[55] O. Luis, "Influence of co-authorship networks in the research impact: Ego network analyses from Microsoft Academic Search," *Journal of Informetrics,* vol. 8, no. 3, pp. 728-737, 2014.

[56] B. C. Heath and T. B. Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems,* Vols. 5,3, pp. 1-22, 2009.

[57] R. Meymandpour and J. G. Davis, "A Semantic Similarity Measure for Linked Data: An Information Content-Based Approach," *Knowledge-Based Systems,* vol. 109, pp. 276-293, October 2016

[58] T. Holst and E. Höfig, "Investigating the Relevance of Linked Open Data Sets with SPARQL Queries," in *37th Annual Computer Software and Applications Conference Workshops,* Berlin Heidelberg, 2013. DOI. 10.1109/COMPSACW.2013.31.

[59] R.-G. Nicolas, T -S. Daniel, H.-V. Enrique, and D. Domingo, "Mining university rankings: Publication output and citation impact as their basis," *Research Evaluation,* vol. 28, no. 3, p. 232–240, 2019.

[60] C Xuanyu, C. Yan and R. L. K.J., "A data analytic approach to quantifying scientific impact," Journal of Informetrics," *Journal of Informetrics,* vol. 10, no. 2, p. 471–484, 2016.

[61] J. E. Hirsch, "An index to quantify an individual's scientific research," in *National Acad Sciences,* 2005.

[62] E. Ferrara and A. Romero, "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,* vol 64, no. 11, p. 2332–2339, 2013.

[63] B. Lutz and H. Daniel-Dieter, "What do we know about the h-index? " *Journal of the American Society for Information Science and Technology,* vol. 59, no. 9, p. 1381–1385, 2007.

[64] A. Daud, J. Li, L. Zhou, and F. Muhammad, "Temporal expert finding through generalized time topic modeling," *Knowledge-Based Systems,* vol. 65, no. 9, pp. 70-81, 2010.

[65] T Amjad, Y. Ding, J. Xu, C. Zhang, J. Tang, M. Song, and A. Daud, "Standing on the shoulders of giants," *Journal of Informetrics,* vol. 11, p. 307–323, 2017.

[66] F Zhang, X. Bai and I Lee, "Author Impact: Evaluations, Predictions, and Challenges," *IEEE Access,* pp. 125-138, April 2019.

[67] W. Dewi Ahmad and A. Abu Bakar, "Classification Models for Higher Learning Scholarship Award Decisions," *Asia-Pacific Journal of Information Technology and Multimedia,* vol. 7, no. 2, p. 131 – 145, December 2018.

[68] X Wu, V. Kumar, and J. R. Quinlan, "Top 10 algorithms in data mining," *Knowledge Information System,* vol. 14, pp. 1-37, 2008.

[69] S. Jamil, The challenge of establishing world-class universities., WashingtonDC: TheWorldBank. ,ISBN978-0-8213-7865-6., 2009.

[70] P Jacek, "Country-specific determinants of world university rankings," *Scientometrics,* vol. 114, p. 1129–1139, 2018.

[71] A. O. Gokcen and B. Melih, "Is Measuring The Knowledge Creation of Universities Possible?: A Review of University Rankings," *Technological Forecasting & Social Change,* vol. 123, pp. 153-160, October 2016.

[72] F Henk and Moed, "A critical comparative analysis of five world university rankings," *Scientometrics,* vol. 110, no. 2, p. 967–990, 2016.

[73] M. Maxime and P J. Alex, "Engineering the Global University Rankings: Gold Standards, Limitations, and Implications," *IEEE Access,* vol. 6, pp. 6765-6776, 2018.

Appendices

A) OpenRank - Top 100 Universities

List of the Top 100 Universities as Ranked by the OpenRank Academic Ranking Methodology (2018)	
Name of the Higher Education Institution (HEI)	**Rank**
University of Cambridge (UC)	1
Massachusetts Institute of Technology (MIT)	2
Stanford University (SU)	3
Carnegie Mellon University (CMU)	4
The University of Illinois at Urbana Champaign (UIUC)	5
Georgia Institute of Technology, Georgia Tech (GIT)	6
Princeton University (PrU)	7
University of Hong Kong (UHK)	8
Purdue University (PU)	9
University of Washington (UW)	10
University of Southern California (USC)	11
University of Toronto (UT)	12
University of Michigan (UM)	13
Pennsylvania State University (PSU)	14
KU Leuven (Katholieke Universiteit Leuven) (KUL)	15
University College London (UCL)	16
National University of Singapore (NUS)	17
Harvard University (HU)	18
University of California at Berkeley (UCB)	19
University of Oxford (UO)	20
University of Sydney (US)	21
The University of Maryland, College Park (UM)	22
University of Florida (UF)	23
The University of Texas at Austin (UTXA)	24
Ohio State University (OSU)	25
Tsinghua University (TU)	26
The University of New South Wales (UNSW)	27
Nanyang Technological University Singapore (NTUS)	28
Australian National University (ANU)	29
National Taiwan University (NTU)	30
City University of Hong Kong (CUHK)	31
University of British Columbia (UBC)	32
The University of Melbourne (UoM)	33
Michigan State University (MSU)	34
Peking University (PkU)	35

List of the Top 100 Universities as Ranked by the OpenRank Academic Ranking Methodology (2018)	
Name of the Higher Education Institution (HEI)	**Rank**
Columbia University (CU)	36
University of Wisconsin Madison (UWM)	37
The University of Colorado at Boulder (UCB)	38
New York University (NYU)	39
Cornell University (CoU)	40
Monash University (MoU)	41
University of Minnesota (UoM)	42
The University of Tokyo (UoT)	43
Swiss Federal Institute of Technology Zurich (SFITZ)	44
The University of Edinburgh (UoE)	45
University of Chicago (UoC)	46
The University of North Carolina, Chapel Hill (UNCH)	47
Uppsala University (UpU)	48
University of Copenhagen (UCH)	49
The University of California at Irvine (UCI)	50
The University of Texas, Dallas (UTD)	51
University of Alberta (UoA)	52
Delft University of Technology (DUT)	53
The University of Manchester (UoM)	54
University of Ghent (Ghent University) (UoG)	55
Mcgill University (McU)	56
University of California at Santa Barbara (UCSB)	57
The University of Auckland (UoA)	58
University of California at San Diego (UCSD)	59
Shanghai Jiao Tong University (SJTU)	60
Northwestern University (NWU)	61
Rice University (RiU)	62
Lund University (LU)	63
University of Pennsylvania (UPen)	64
Fudan University (FdU)	65
University of Durham (Durham University) (UoD)	66
University of Helsinki (UoH)	67
Duke University (DkU)	68
The University of Western Australia (UWA)	69
University of Glasgow (UGl)	70
Aarhus University (AhU)	71
The University of California at Davis (UCD)	72
Boston University (BoU)	73

List of the Top 100 Universities as Ranked by the OpenRank Academic Ranking Methodology (2018)	
Name of the Higher Education Institution (HEI)	**Rank**
University of Pittsburgh (UPb)	74
Osaka University (OsU	75
Imperial College London (The Imperial College of Science, Technology, and Medicine) (ICL)	76
King's College London (KCL)	77
The University of Sheffield (UoSh)	78
Utrecht University (UtU)	79
University of Oslo (UO)	80
University of Leeds (ULd)	81
Johns Hopkins University (JHU)	82
Lomonosov Moscow State University (LMSU)	83
University of Zurich (UZ)	84
University of Southampton (USh)	85
The University of California at Los Angeles (UCLA)	86
University of Basel (UBs)	87
Hong Kong University of Science and Technology (HUST)	88
University of Groningen (UGr)	89
Korea University (KoU)	90
University of Warwick (UWk)	91
The University of Queensland (UQl)	92
Technical University of Munich (TUM)	93
California Institute of Technology, Caltech (CITC)	94
University of Amsterdam (UAm)	95
The Chinese University of Hong Kong (CUH)	96
Heidelberg University (Ruprecht-Karls-Universitat Heidelberg) (HdU)	97
University of Bristol (UBr)	98
Brown University (BrU)	99
Humboldt University of Berlin (Humboldt-Universitat Zu Berlin) (HUB)	100

B) ResRank - Top 50 Institutions

List of the top 50 institutions w.r.t to the ResRank methodology in the year 2020. The institutions have a significant number of Influential Researchers in the Computer Science discipline.

SN	Institution	RC	SN	Institution	RC
1	Stanford University	67	26	University of Massachusetts Amherst	25
2	Carnegie Mellon University	66	27	University of Maryland	24
3	Massachusetts Institute of Technology	59	28	Imperial College London	23
4	University of California, Berkeley	51	29	UCL	21
5	**Microsoft Research**	**48**	30	The University of British Columbia	21
6	**Google LLC**	**47**	31	University of Cambridge	20
7	**The University of Illinois at Urbana-Champaign**	**47**	32	City University of Hong Kong	20
8	**The University of Texas at Austin**	**38**	33	Hebrew University of Jerusalem	20
9	**University of Southern California**	**38**	34	The Electrical and Computer Engineering Department	20
10	University of Toronto	37	35	New York University	19
11	**Ecole Polytechnique Fédérale de Lausanne**	**35**	36	Purdue University	19
12	University of California, Los Angeles	34	37	Rheinisch-Westfälische Technische Hochschule Aachen	19
13	Georgia Institute of Technology	34	38	Brown University	19
14	**University of Michigan, Ann Arbor**	**34**	39	University of California, Irvine	19
15	University of Washington, Seattle	30	40	Harvard University	18
16	ETH Zürich	30	41	Aalto University	18
17	Cornell University	29	42	Chinese University of Hong Kong	18
18	University of Oxford	29	43	Duke University	18
19	Princeton University	28	44	Hong Kong University of Science and Technology	18
20	**Technion - Israel Institute of Technology**	**28**	45	Nokia Bell Labs	18
21	Tel Aviv University	28	46	Northeastern University	18
22	University of Waterloo	28	47	Technical University of Munich	18
23	University of California, San Diego	27	48	University of California, Davis	18
24	Columbia University in the City of New York	26	49	National University of Singapore	17
25	University of California, Santa Barbara	25	50	Arizona State University	17

- **SN** = Serial Number **RC** = Researchers' Count
- A name in the bold font indicates that the institution is among the top-20 (w.r.t the ResRank Methodology) but it is not among the top-20 w.r.t the Grand Average Rank (GAR).

C) The Scopus dataset - Statistics

Table A.0.1 Count of the researchers in the Scopus repository for different domains and fields

Domain	Field	Authors' Count
Applied Sciences	Agriculture, Fisheries & Forestry	256740
	Built Environment & Design	45286
	Enabling & Strategic Technologies	624691
	Engineering	601302
	Information & Communication Technologies	**470,403**
Arts & Humanities	Communication & Textual Studies	31148
	Historical Studies	33141
	Philosophy & Theology	19041
	Visual & Performing Arts	4757
Economic & Social Sciences	Economics & Business	134945
	Social Sciences	153733
Health Sciences	Biomedical Research	616344
	Clinical Medicine	1996452
	Psychology & Cognitive Sciences	95652
	Public Health & Health Services	161942
Natural Sciences	Biology	267938
	Chemistry	494121
	Earth & Environmental Sciences	237406
	Mathematics & Statistics	95390
	Physics & Astronomy	646860
	Total	**6,987,292**

Table A.0.2 World Top 2-Percent (WT2P) researchers as ranked by the "Composite Indicator"

In November 2020, PA Ioannidis et al. from the US-based Stanford University released a dataset of the top 2 percent of the most-cited scientists in various scientific disciplines. The dataset is a derivative of the Scopus researchers' profiles, containing a record of 1,59,683 scientists who are ranked using a newly developed research evaluation metric termed as "Composite Indicator".	**Statistics**
Authors' Count	159,684
Publication Span	1960-2019
Citation Span	1996-2019
H-Index	Up to Dec. 2019
Distinct HEIs in the author's affiliation	2396
Distinct Countries in the author's affiliation	21

Table A.0 3 Count of the researchers in the eight sub-disciplines of ICT

Field (Sub-Discipline)	Researchers' Count	
	When the author's second field is NOT restricted	When the author's second field is restricted to the 8 sub-disciplines.
Artificial Intelligence & Image Processing	215114	1939
Computation Theory & Mathematics	16572	240
Computer Hardware & Architecture	17080	163
Distributed Computing	9666	164
Information Systems	16581	145
Medical Informatics	13000	30
Networking & Telecommunications	161179	1262
Software Engineering	21211	329
Total	**470,403**	**4,272**

ICT = Information & Communication Technologies