International Journal of Data Science and Analytics (2025) 20:3933–3950
https://doi.org/10.1007/s41060-024-00702-x

REGULAR PAPER

Investigating scholarly indices and their contribution to recognition patterns among awarded and non-awarded researchers

Bilal Ahmed[1,2] **· Li Wang**[1] **· Waqar Hussain**[2] **· Ghulam Mustafa**[3] **· Muhammad Tanvir Afzal**[3]

Received: 18 August 2024 / Accepted: 11 December 2024 / Published online: 15 January 2025

Abstract

The systematic evaluation and ranking of researchers in academia stand out as a challenging and tedious task. This task facilitates policymakers in making informed decisions regarding faculty promotions, awards and providing grants by identifying deserving individuals. Researchers have put forth a range of parameters for identification of the most impactful scholars. These parameters include the count of publications and citations, along with the h-index. However, considering the associated shortcomings of the h-index, the researchers have introduced several variations of the h-index. Fictitious scenarios and datasets will be utilized to evaluate the practical and real-world effectiveness of the proposed metrics. There is a gap that requires a thorough evaluation of the specific behavior of each index. The current study emphasizes the assessment of various indices using a comprehensive dataset that includes both award-winning and non-award-winning researchers. During the evaluation of each index, the retrieval of awardees was examined within the top 10% to top 100% of the ordered list for each respective index. In addition, we investigate the correlation among several indices and awarding societies. Our findings suggest that the A-index and normalized h-index can identify over 60% of award-winning researchers.

Keywords Researcher ranking · Expert finding · Bibliometric measures · Variants of h index · Research evaluation

1 Introduction

In academia, a tremendous amount of scholarly articles are generated by researchers on a daily basis [1–3]. As the volume of research articles increases exponentially, it is important to build a mechanism for the evaluation or ranking of researchers [4, 5]. The evaluation of researchers' contributions has emerged as a critical aspect of the scientific endeavor [6, 7]. Rankings provide a valuable tool for the scientific community in making crucial decisions, such as selecting candidates for scientific awards, allocating scholarships and grants, and determining tenure-track positions and promotions [8–11]. Rankings also assist in selecting editors and reviewers for journals and conferences, as well as identifying prominent experts in specific domains. Furthermore, rankings aid students in identifying suitable research supervisors for their doctoral studies. This highlights the extensive and longstanding tradition of utilizing researcher rankings [12, 13].

Publication count has traditionally been a conventional metric for evaluating a researcher's output. Although this approach works well, it also has certain drawbacks. For example, an author who has authored only a limited num-

B. Ahmed, L. Wang, W. Hussain, G. Mustafa and M. Afzal have contributed equally to this work.

✉ Bilal Ahmed
bilal.ahmed.raza@uokajk.edu.pk

Li Wang
wangli@tyut.edu.cn

Waqar Hussain
waqar.hussain@uokajk.edu.pk

Ghulam Mustafa
ghulam.mustafa.ssc@stmu.edu.pk

Muhammad Tanvir Afzal
dean.foc@stmu.edu.pk

1 Department of Computer Science, Taiyuan University of Technology, Taiyuan, Shanxi, China

2 Department of Data Science and Artificial Intelligence, University of Kotli AJ&K, Kotli 11100, Azad Kashmir, Pakistan

3 Department of Computing, Shifa Tameer-e-Millat University, Islamabad 44000, Pakistan

ber of papers, all of which possess significant influence and impact, may surpass another author who has produced a larger bulk of research publications but within journals or conferences with lower or no impact factors. This metric does not accurately reflect the scholar's significance or scientific influence. It only serves as an indicator of publication quantity.

While a substantial citation count [14] may be indicative of influence and recognition within the scientific community, it's not always a dependable indicator of a researcher's work or long-term impact. For example, numerous research groups solely reference their own research articles in order to increase the citations of their research articles.

To address the limitations of relying solely on publication count or citation frequency, Jorge E. Hirsch proposed a novel approach in 2005, known as the h-index, to assess the scholarly influence of researchers [15]. The Hirsch index assesses a researcher's scientific influence through a singular numerical measure.

However, the scientific community has recognized several shortcomings of the h-index. For instance, the h-index does not account for self-citations by researchers, which can artificially inflate impact measurements. Additionally, the h-index has limitations in that it cannot decrease over time, even if a researcher's recent work receives fewer citations. Another limitation is that the h-index may show bias toward researchers who publish in high-impact journals or have a higher volume of publication, as it inherently favors quantity and citations over the quality or significance of individual papers [16]. To address these limitations, various novel indicators have been proposed. A recent study [17] published in 2023 suggests that over eighty different indicators have been proposed to measure the research impact of scholars. A new area of research has emerged with the introduction of the h-index. As of now, the h-index has been cited 15,550 times, and researchers have proposed several modifications of the h-index into various parameters. Moreover, over 27,000 research articles have been published wherein researchers evaluate or propose new indices [18]. Despite the continuous introduction of new indices, a majority of these have been tested in fictional or hypothetical scenarios, specific datasets, or use cases. Therefore, undertaking a comprehensive evaluation of the significance of these quantitative parameters presents a significant challenge.

The current study investigates the function of several h-type indices, including A index, E index, G index, M index, K index, Normalized h index, P index, Pi index, q2 index, R index, Real h index, Rm index, W index and Weighted h index using a comprehensive dataset of researchers of a particular domain. Our main focus in this study is to find the potential metrics that can be used while evaluating researchers in the mathematics domain. Our study utilizes a dataset of 1050 researchers, comprising 525 awardees and 525 non-awardees, spanning three decades from 1990 to 2023. This data set is also published in our recent study and can be downloaded from the link.[1] The information about the awardees researchers is gathered from various mathematical societies that have recognized and honored researchers in the field. Some well-known societies in the domain of mathematics were employed as a yardstick in our analysis as follows.

AMS "American Mathematical Society',[2] IMU "International Mathematical Union",[3] LMS "London Mathematical Society",[4] NASL "'Norwegian Academy of Science and Letters"[5].

The subsequent sections of this paper are structured as follows: the "Literature Review" section, we provide a concise overview of recent studies related to ranking indices. The "Methodology" section delineates our proposed approach for evaluating the indices. In the "Results and Discussions" section, we present and explore the outcomes of our research. The "Conclusion" section will summarize the key findings and their implications. In addition, this section also discussed the limitation of study and will delve into exploring some future directions.

2 Literature

Over the past decade, an enormous amount of research has been carried out every day by many researchers. However, with the rapid growth in research output, there is an extreme need for an efficient mechanism to evaluate and rank researchers based on their scientific contributions [19].

1 https://github.com/bilalahmedraza/project.

2 https://www.ams.org/home/page.

3 https://www.mathunion.org/.

4 https://www.lms.ac.uk/.

5 https://dnva.no/norwegian-academy-science-and-letters.

An effective ranking system will empower the community to make informed decisions in diverse contexts, such as: who should be hired or promoted for a specific position. The government uses this ranking system for various categorization purposes, such as allocating research funding to top institutions. A comprehensive ranking system plays a significant role in helping international students to choose the best universities and research supervisors for their postgraduate studies. In addition, a fair ranking system will help journal, and conference editors identify the best reviewers from around the world [20].

Diverse methods have been implemented within the scientific community to evaluate researchers across various fields. Each approach adopts a distinct methodology for ranking researchers based on their respective contributions. In 2023, researchers Lathabai and Prabhakaran introduced an index called the Contextual Index [21] for ranking the most influential researchers. They conducted a comparative study, evaluating the performance of this index against the well-known h-index and g-index. The dataset for metric evaluation was derived from the profiles of 100 researchers across various domains, all with citation counts ranging from 100 to 3000. Their analysis led to the conclusion that the Contextual Index outperforms the other indices.

Researchers Sharma and Uddin proposed a novel index named the Kz Index for ranking scholars. They contend that the proposed index effectively ranks researchers in comparison to traditional measures such as the g-index and h-index. The experiment involved a dataset comprising 376 researchers from different domains at Monash University [22].

In [8] researchers assessed author count and publication age-based indices on the same dataset of researcher. In 2022, researchers proposed an index named the hc Index for ranking researchers. They assert that the proposed index demonstrates effective performance, especially for lower-ranked researchers when compared to the h-index. The performance of the proposed index was assessed using a dataset encompassing the publication records of 385 Monash University authors. Basu et al. in [23] conducted an evaluation of the h-index for the purpose of ranking universities in India. In 2015, Knight [24] collected data from 37 neurosurgical departments in the UK and Ireland. He ranked these departments solely based on the h-index across different time spans, without applying any specific threshold criteria. In 2014, researchers Rubem et al. [25] conducted a comparative analysis of various bibliometric indicators, encompassing the h-index and fifteen of its variants. Their evaluation was conducted on a dataset comprising six individual scientists aggregated into seven distinct groups. In 2008, researcher Schreiber [26] introduced the hm-index, which takes into account the number of co-authors when ranking researchers. The experiments for this index were carried out using a database of physicists, which included data from conference and journal publications. In 2007, a team of scholars conducted an assessment of the g-index in comparison to the h-index, a-index, and r-index. The dataset utilized for evaluating these indices consisted of 26 physicists. Their evaluation results demonstrated that the g-index is a more suitable metric than the h-index for evaluating the scientific impact of a researcher. In 2006, researcher Kosmulski [24] computed the correlation among the h-index and the h2 index. These experiments were conducted using data from 19 professors affiliated with the Department of Chemistry at a university in Poland.

Table below compiles a set of relevant studies in which researchers evaluate and introduce various parameters across different domains for ranking researchers. Upon scrutinizing these studies, we observed a common trend wherein the evaluation encompasses a relatively small number of parameters and is conducted using datasets comprising a limited number of researchers from various domains.

D3D3D3 Author	Title of Paper	Dataset	Indices	Result	Limitation
[27]	"Four Limitations of the h-Index for Evaluating the Research Productivity and Impact of Individual Authors"	Google Scholar data on 12 Nobel laureates across four scientific fields	H index, hi index	Hi index is better than h index for ranking researchers	Limited dataset, very few indices used
[28]	"The h-index is no longer an effective measure of scientific reputation"	4000 authors data from four different fields (biology, physics, computer science, economics)	fractional h index, H index	Fractional h index is superior to h index for ranking researchers	Few parameters used
[29]	"Comparison of researchers' impact indices"	Computer Science Dataset (Arnet Miner dataset)	H index, complete h, k index	K index performs better than other indices for ranking researchers	Very few parameters used
[30]	"Assessment of the h-Index and Its Qualitative and Quantitative Variations in Neuroscience"	Number of researchers: 547 Publications: 96,327 Citations: 5,850,966 Awardees: 48	h-index, m-Quotient, e-index, r-index, hg-index, f-index	r index and hg index are better for ranking researchers	Very few parameters used
[31]	"Assessment of the h-Index, Its Variants, and Extensions Considering Publication Age and Citation Intensity in Civil Engineering"	Civil Engineering: Authors: 36,921 Publications: 20,307 Citations: 2,184,638	T index, A index, H index, F index, ht index, ar index, Raw h, q2 index, Wu index, Hc index, Maxprod	In the civil engineering domain, Wu index is better for researcher ranking	Very few parameters used
[32]	"Ranking of universities in India"	Universities: 39 Papers: 58,781 Citations: 590,227 Faculty per university	h-index	Central universities in India ranked without a benchmark	Limited features used No benchmark
[33]	"Assessment of neurosurgical researchers in United Kingdom and Ireland"	37 neurosurgical departments' data from UK and Ireland	h-index (2004-2014) and (2012-2014)	Neurosurgical departments ranked without a benchmark	Limited features used No benchmark

Previous studies have attempted to evaluate the importance of ranking parameters by utilizing hypothetical or fictional scenarios. These studies lack a comprehensive benchmark dataset to validate their findings. However, the performance of each index can only be attained through empirical investigations that utilize extensive and diverse dataset of specific field. For this reason, our research study employed a comprehensive dataset of authors of three decades from 1990 to 2023, specifically for the domain of mathematics.

3 Methodology

The scientific community has proposed several approaches for ranking the employees in organizations. After critically analyzing, it has been determined that most employees are ranked based on publication counts, citation counts, or the h-type indices [17], [34]. Figure 1 demonstrates the block diagram that outlines the suggested methodology.

3.1 Data gathering

Gathering data for a specific domain like mathematics needs input from domain experts due to the field's diverse branches. Mathematics encompasses a diverse array of subfields, including algebra, calculus, probability theory, geometry, number theory, and many more. Each subfield needs specialized knowledge for accurate data collection. The Mathematics Subject Classification (MSC)[6]system categorizes these branches hierarchically, aiding researchers in navigation.
With the help of domain expert, we identify the list of terms from MSC and then manually gather data of active researchers from Google Scholar. Several sources are available to gather data on authors' research activities, including publications, citations, co-author networks, and other relevant metrics. Some well-known sources include MathSciNet, Zbmath, WoS, and Google Scholar. In our research study, we faced challenges related to limited coverage and access issues with some of the sources we initially considered. For instance, certain databases or repositories may impose access restrictions, necessitate a subscription or membership, or exhibit limited coverage of specific disciplines or publication type. To address these challenges, we explored alternative sources and eventually decided to obtain the dataset from Google Scholar. In addition, studies have shown that Google Scholar is the best choice among available platforms for collecting researchers' metadata [35]. Figure 2 shows different document typologies that are only available in Google Scholar digital library.

[6] https://zbmath.org/classification/

The selection of this particular resource was guided by several key factors, including its comprehensive coverage of scholarly publications across a wide range of disciplines, its global accessibility for researchers, and its ability to retrieve both open-access and subscription-based publications. Multiple research studies have compared the coverage of Web of Science (WoS) and Google Scholar, consistently demonstrating that Google Scholar's growth exceeds that of WoS by 13%. Moreover, the citations within Google Scholar exhibit an average monthly growth rate of 1.5%27. Furthermore, Google Scholar is a dynamic and continuously evolving platform that incorporates the latest data on a weekly basis, guaranteeing the timeliness and relevance of its information [35]. We use Publish and Perish platform to retrieve data from Google Scholar. Publish and Perish employs an advanced algorithm to extract both primary data and related metadata of authors from Google Scholar as shown in Fig. 3.

After collecting this metadata of researchers, we preprocess the collected data, validating its alignment with mathematics while also executing author disambiguation and eliminating any irrelevant noise. Several steps have been taken in data preprocessing. The workflow of preprocessing is shown in Fig. 4

3.1.1 Removal of irrelevant characters

A total of 204649 publications were gathered from Google Scholar. For inspecting each publication, we developed a Python script to analyzed each publication's title to identify the presence of special symbols such as ($, %, # and others) that could cause issues in data analysis and interpretation. Publications with titles dominated by these symbols and lacking substantive content were removed using predefined heuristics during the preprocessing phase. This ensured that the remaining publications were cleaner and more reliable for further analysis.

3.1.2 Research papers verification

The initial dataset consisted of 205,649 research publications spanning various fields. To assess its relevance to Mathematics, a random sample of 1700 publications was drawn using the Raosoft application. Domain experts reviewed this sample and found that 92% of the publications were related to Mathematics, indicating a high degree of relevance. Following this, the dataset underwent a filtration process based on publication venues. Each publication was checked against a curated list of recognized Mathematics journals and conferences. Publications appearing in these venues-195,000 in total-were directly included in the final dataset. For the 9867 publications from non-Mathematics venues, each title was reviewed by domain experts to confirm relevance. During this review, 1169 publications were identified as unrelated

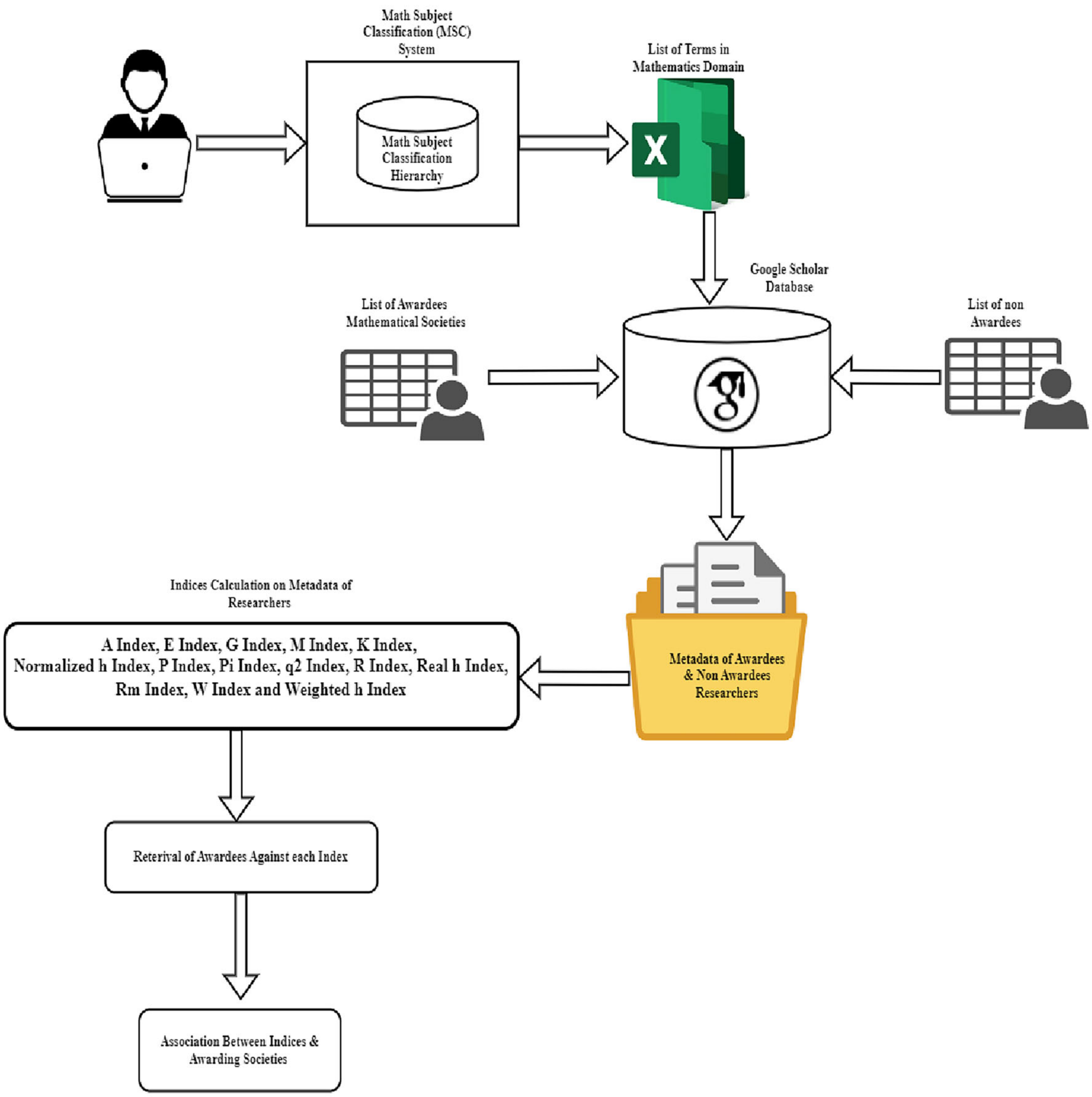

Fig. 1 Workflow of the proposed methodology

to Mathematics and were removed. Consequently, the final dataset comprised 204,649 publications, validated as relevant to the field of Mathematics.

3.1.3 Name disambiguation

Furthermore, as part of the data preprocessing procedures, author disambiguation is conducted. Several studies have recognized the necessity for author disambiguation in Google Scholar results. This procedure entails the meticulous identification and removal of duplicate entries and the resolution of any inconsistencies in authors' first or last names.

Table 1 presents the final dataset description for the evaluation, including the number of publications and citations for several researchers.

3.2 Computation of indices

This section presents both the calculation method and a brief introduction to these indices. After gathering the data, indices were calculated using the researchers' collected metadata. A Python utility was used to compute all the indices. Our aim in presenting this information is to provide the reader with a comprehensive understanding of the various h-index variants employed in this study.

3.2.1 A index

Researcher Burrell [36] introduced a new index known as the A-index. The A-index of a researcher is determined as the mean number of citations received by their h-core articles as

Fig. 2 Document typologies in google scholar

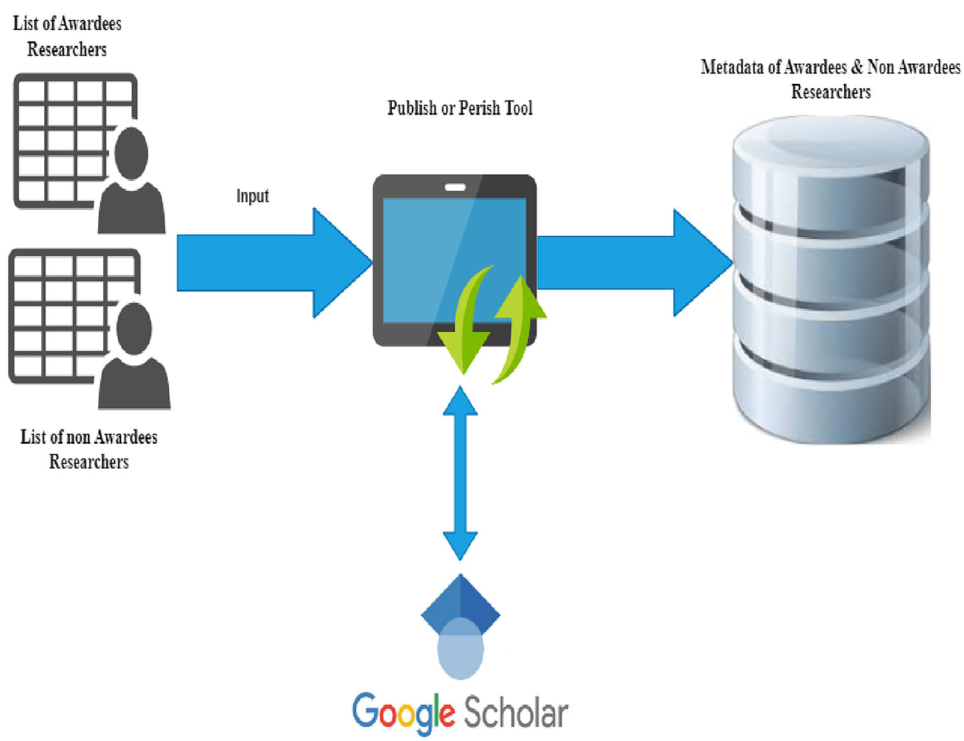

Fig. 3 Metadata collection workflow

Table 1 Metadata of researchers for Indices evaluation

D3D3D3 Researchers	Count of features
Total number of award winners	Five hundred and twenty-five (525)
Total number of non-award winners	Five hundred and twenty-five (525)
Total number of citations of researchers	14,370,049
Total number of publications of researchers	204,649

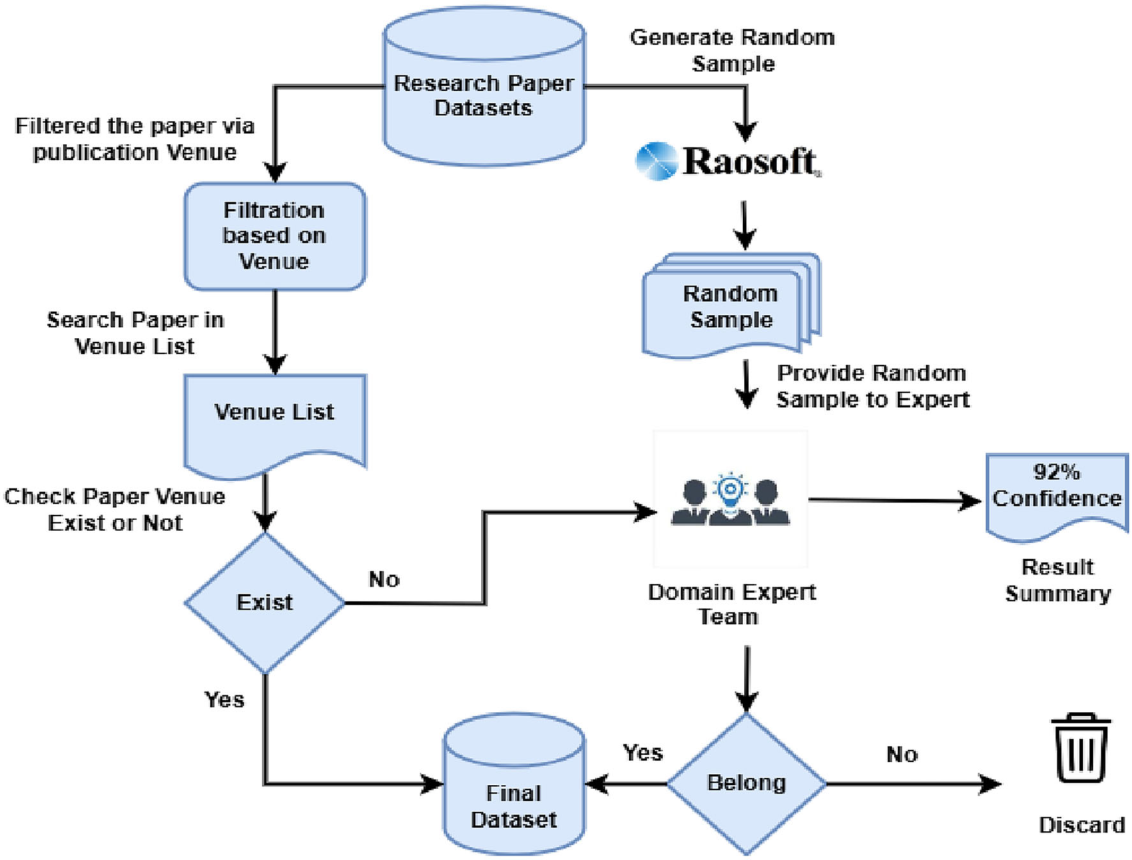

Fig. 4 Research paper verification

expressed in the formula below

$$A = \frac{\sum_{i=1}^{h} c_i}{h} \tag{1}$$

3.2.2 E index

Zhang introduced a novel index named E index [37]. This index represents the citations within the h-core that are not accounted for by the h-index. If e is equal to zero, the h-index does not lose any information since it fully describes the citation information for papers in the h-core. The purpose of this index is to capture the surplus citations that are disregarded by the h-index. Mathematically, it can be expressed as follows

$$e = \sqrt{\sum_{i=1}^{h} (c_i - h)} \tag{2}$$

3.2.3 g index

Egghe and Rousseau [38] proposed g index. The g index is defined as the maximum number g of highly cited articles for which the top g articles have accumulated or more citations. Mathematically, it can be represented as

$$\sum_{i=1}^{g} c_i \geq g^2 \tag{3}$$

3.2.4 K dash index

Chen et al. [39] proposed a new index to address the divide by zero obstacle of k-index called K/ index. Mathematically the K/ index is computed as follows

$$K' = \sqrt{\frac{h \cdot C}{N}} \tag{4}$$

3.2.5 M index

Bornmann proposed M index [40]. The m-index is defined as the median number of citations of the h-core published papers. Unlike measures such as the mean or square root of the citation count, the median is less affected by extreme values and provides a more representative estimate of the typical citation count for a scholar's articles. By considering the distribution of citations within a scholar's h-core, the m-index offers a more precise assessment of their research impact

$$m = \frac{h}{n} \tag{5}$$

3.2.6 Normalized h index

Sidiropoulos et al. [41] introduced the Normalized h-index as a metric to tackle the fairness concern when comparing the impact of scholars with varying numbers of publications. The Normalized h-index resolves this issue by normalizing the h-index in relation to the number of publications. The index is calculated by dividing the h-index of a scholar by the square root of their total number of publications. This normalization ensures that the impact of a scholar is equitably compared, irrespective of the number of publications they possess. Mathematically, it can be expressed as follows:

$$h_{\text{norm}} = \frac{h}{n} \tag{6}$$

3.2.7 P index

A novel performance measure, the p-index, is introduced by [42]. The p-index aims to achieve an optimal balance between the total number of citations (C) and the mean citation rate (C/P). Mathematically, it can be described as follows

$$P - index = \left(\frac{C^2}{p}\right)^{\frac{1}{3}} \tag{7}$$

3.2.8 Q^2 index

Researchers introduced a new index the Q^2 index [43] similar to the Hg-index, introduces a dependency on the M-index in addition to the H-index. The Q^2 index is computed by extracting the square root of the product of the H-index and M-index. Mathematically, it can be represented as follows

$$Q^2 = \sqrt{h * m} \tag{8}$$

3.2.9 R index

Jin et al. [44] proposed the R-index of researcher and is defined as the square root of the cumulative sum of the citation counts of their h-core articles. Mathematically, this index can be expressed as follows

$$R - index = \sqrt{\sum_{p=1}^{h} Cit_p} \tag{9}$$

3.2.10 Real h index

Guns and Rousseau [45] proposed a new index called real h-index and it can be computed as

$$h_r = \frac{(h+1)cit_h - h.cit_h + 1}{1 - cit_{h+1} + cit_h} \tag{10}$$

3.2.11 Rm index

Panaretos [44] proposed the Rm-index and is defined as the square root of the summation of the square roots of the citation counts of the h-core articles. In essence, it is a weighted measure that considers both the size of the h-core and the distribution of citations among the articles. Mathematically, the Rm-index can be computed as follows

$$R_m = \sqrt{\sum_{k=1}^{h} \text{Cit}_k^{\frac{1}{2}}} \tag{11}$$

3.2.12 W index

Researchers proposed the w index [46]. If an author has at least 10w citations in each of their papers, while some other papers have less than or equal to 10(w+1) citations, then it is referred to as the researcher's w-index. The value of the w-index is equal to w index in this case

3.2.13 Weighted h index

Researchers proposed weighted h index [38] to capture the sensitivity of a scholar to variations in publication and citation count. Mathematically, the weighted rank of an article can be expressed as follows

$$R_w(k) = \frac{\sum_{p=1}^{k} Cit_p}{h} \tag{12}$$

Following the calculation of each metric, distinct ranking lists were generated to reflect the unique standings associated with each index. Consequently, an individualized ranking list for authors was created based on each metric, providing a comparative overview of author performance across different indices. These results are visually represented in Fig. 5.
Figure 5 above displays the names of the authors along with their corresponding index values. The final label, 'Class,' indicates the binary values of 1 and 0, signifying whether an author has received an award (1) or not (0). Due to the extensive number of indices calculated, only a subset is presented in the figure for clarity and readability.

Fig. 5 Researchers score against each index

Author Name	E Index	G Index	H Index	Q2 Index	Real h index	W Index	Weighted H Index	Class
A. Corti	6,400	39	39	72	-1	16	76	1
A. Spielman	23,056	186	116	150	59	31	160	1
Brooke Shipley Hrushovski	2,992	71	37	56	20	12	56	0
DL Powers	840	40	23	30	8	6	32	1
H Stehfest	5,415	74	10	26	11	5	70	0
J.B. Mcleod	3,593	77	40	53	11	10	59	0
James Newton	124,715	440	206	272	42	61	347	0
Jean-christophe Yoccoz	674	21	11	25	6	5	26	0
John Huerta	61,985	132	82	214	21	44	238	1
Keisuke Hara Zilber	25	6	3	7	2	1	6	1
Kiyosi It	12,720	87	30	75	3	16	100	1
M. Owen	63,188	118	118	219	-1	45	232	1
M.J.D. Powell	35,223	161	63	109	33	27	179	0
RP Agarwal	4,170	86	45	60	16	14	67	0
SY Shaw	1,155	45	24	35	25	7	35	0
U Koumba	37	8	4	7	3	1	6	1

4 Results and discussions

This section presents a comprehensive and detailed description of the analysis results. First, we calculated the metrics for each researchers, and we then proceeded to analyze the proportion of award recipients associated with each metric. Our analysis investigated the frequency of awardee presence across different metric categories, encompassing the top 10%, top 50%, and top 100% of the generated ranked lists.

The notable prevalence of the A index, m index, and normalized h index in identifying a substantial proportion (around 70%) of award recipients at top 10% is consistent with prior literature emphasizing the effectiveness of these indices in assessing the performance of researchers. The A index, renowned for its consideration of both the quantity and impact of publications, is likely identifying researchers with enduring and impactful contributions, frequently reflective of accomplishments deserving of awards.

Likewise, the m index, which takes into consideration both productivity and citation impact, showcases its efficacy in pinpointing researchers with a significant influence, a trait frequently linked to awards and recognition. These indices, by integrating both productivity and citation metrics, offer a thorough perspective on a researcher's impact within the scholarly community.

On the flip side, the lower retrieval rate (10%) noted for the k dash index and rm index may be ascribed to the particular emphases of these indices. The k dash index places significance on the novelty and impact of a researcher's work, potentially resulting in a more limited identification of awardees. Meanwhile, the rm index, centered on recent research output, might overlook individuals with a longer history of impactful contributions.

The disparity in retrieval percentages observed between the A-index and the g-index within the top 100% of the ranked list highlights the inherent differences in the underlying metrics and methodologies employed by these indices.

The A index, amalgamating both the quantity and impact of publications, showcases its effectiveness in pinpointing prolific researchers with a substantial body of influential work. Awardees, frequently acknowledged for sustained excellence, are more inclined to possess a significant presence in terms of both publication volume and citation impact, thereby contributing to the observed 60% retrieval rate.

In contrast, the g-index, while also considering both the quantity and impact of publications, places a specific emphasis on highly cited works. This characteristic may lead to a more discerning identification of researchers whose work has had a pronounced and widespread impact on the scholarly community. Hence, the observed 17% retrieval rate for award winners by the g index suggests that this index is more selective in recognizing researchers whose contributions have resulted in substantial citation counts, potentially at the expense of some prolific but less-cited individuals.

These findings emphasize the significance of comprehending the distinct emphasis and criteria of each index when interpreting their outcomes. The g index, with its concentration on highly-cited papers, might prove more adept at identifying researchers whose work has exerted a transformative influence on the field, even if their publication record is not as extensive as those captured by the A index.

4.0.1 Percentage of award recipients in the top 10 percentile of the ranked list

Figure 6 reveals that the A-index, normalized h-index, and M-index effectively identifies 70% of the awardees. Conversely, the K-dash and Rm indices demonstrate poor performance, identifying only up to 10% of the awardees in the top 10% of the ranked list.

4.0.2 Percentage of award recipients in the top 50 percentile of the ranked list

Figure 7 shows that the normalized h-index and A-index effectively identify over 60% of the awardees within the top 50% of the ranked list. In contrast, the G-index performs poorly, retrieving only 10% of the awardees.

4.0.3 Percentage of award recipients in the top 100 percentile of the ranked list

Drawing upon the data presented in Fig. 8 it becomes evident that the A-index and normalized h-index successfully identify 60% of the awardees within the top 100% of the ranked list. In contrast, the G-index exhibits significantly lower performance, only managing to identify 17% of the awardees.

The illustration presented below in Fig. 9 provides a visual depiction of the combined outcomes across diverse percentile ranges. This graphical representation delivers a holistic perspective on the results achieved within distinct percentiles, offering a nuanced comprehension of the distribution of data and trends across the entire range.

4.1 Relationship between awarding societies and indices

In this section we perform an analysis that which awarding society is depending on which index. We have already mentioned in the figure that our benchmark data set consists of 525 award winner and an equal number of non-award winner. According to a prevailing myth that award winners possess a robust research background, as evidenced by their substantial publication and citation records. Consequently, it has been commonly anticipated that all award winner would consistently rank within the top 10% of researchers when sorted according to these indices. But recent analyses have challenged this assumption, highlighting cases where certain award recipients do not meet this expected criterion. The below Figs. 10, 11 and 12 shows the result of indices associated with different societies.

Through an analysis of the relationship between awarding societies and indices based on author count, we have made the following observations.

4.1.1 Results for American mathematical society (AMS) at top 10%, 50% and 100%

- Within the uppermost 10% of the outcomes, the Pi index and k-dash index exhibited superior performance compared to other indices, successfully retrieving 100 percent of the awardees in the ranked list. Additionally, the G index captured 50 percent of the award winner in the ranked list. However, the P index demonstrated poor performance specifically within the AMS society as shown in figure.
- Within the uppermost 50% of the outcomes, the G index exhibited superior performance in comparison with other indices, retrieving 80 percent of the award winners in the ranked list. On the other hand, the Q2 index showed a performance level of 48 percent as shown in Fig. 10.
- Within the top 100% of the results, the G index displayed a performance level of 70 percent, whereas the Real h index exhibited a performance level of 45 percent, as depicted in Fig. 11.

4.1.2 Results for international mathematical union (IMU) at top 10%, 50% and 100%

- Within the top 10% of results, the Rm index demonstrated a perfect performance level of 100 percent, indicating a successful retrieval of all awardees. Meanwhile, the Q2 index exhibited a performance level of 70 percent, capturing a significant portion of the awardees in the ranked list as shown in figure.
- Within the top 50% of results, the Pi index and weighted h index displayed a performance level of 25 percent, indicating the retrieval of a quarter of the awardees in the ranked list. On the other hand, the G index, Rm index, and R index demonstrated a performance level of 20 percent, capturing a fifth of the award winners in the ranked list as shown in figure.
- In the top 100% of results, the R index exhibited a performance level of 17 percent, retrieving a relatively small proportion of the award winners in the ranked list. Conversely, the G index demonstrated a much lower performance level of only 5 percent, indicating a limited ability to identify awardees as shown in figure.

4.1.3 Results for LMS at top 10%, 50% and 100%

- Within the top 10% of the results, the Normalized h index displayed a perfect performance level of 100 per-

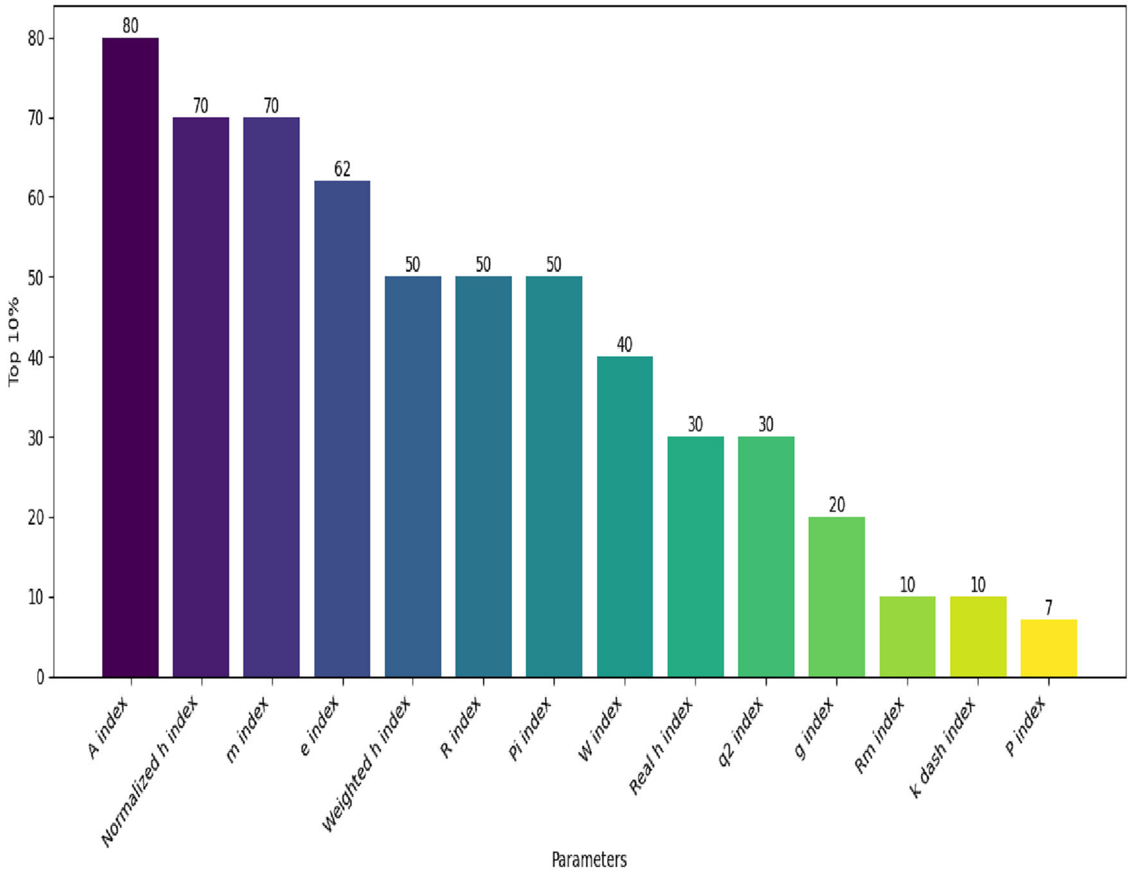

Fig. 6 Awardees percentage at top 10%

cent, successfully retrieving all awardees. Conversely, the Q2 index exhibited a performance level of 70 percent, capturing a significant portion of the award winners in the ranked list. Notably, the lowest performance was observed with the A index.

- Within the top 50% of results, the K dash index demonstrated a performance level of 75 percent, indicating a substantial retrieval of award winners in the ranked list. Conversely, the W index, M index, P index, and Q2 index exhibited a performance level of 45 percent, capturing a smaller proportion of the awardees in the ranked list.
- Within the top 100% of results, both the Q2 index and the Normalized H index performed well, demonstrating a level of performance of 50 percent. However, the G index exhibited a significantly lower performance level of only 5 percent, indicating limited success in identifying awardees.

4.1.4 Results for Norwegian academy of science and letters (NASL) at top 10%, 50%, 100%

- Within the top 10% of results, the Real h index demonstrated a performance level of 30 percent, indicating the retrieval of a relatively moderate proportion of award winners in the ranked list. Conversely, both the A index and E index exhibited equal performances, reaching up to 10 percent, suggesting a lower success rate in identifying awardees.
- Within the top 50% of results, the Real h index displayed a performance level of 17 percent, indicating a moderate success in retrieving awardees within the ranked list. Conversely, the Pi index, Weighted h index, A index, R index, and E index exhibited equal performances, reaching up to 5 percent, suggesting a relatively lower success rate in identifying awardees.
- Within the top 100% of results, both the G index and E index exhibited a performance of 15 percent, indicating a moderate success in retrieving awardees within the ranked list. On the other hand, the K dash index and Q2 index showed very low performances, reaching up to only 4 percent, suggesting a minimal ability to identify awardees.

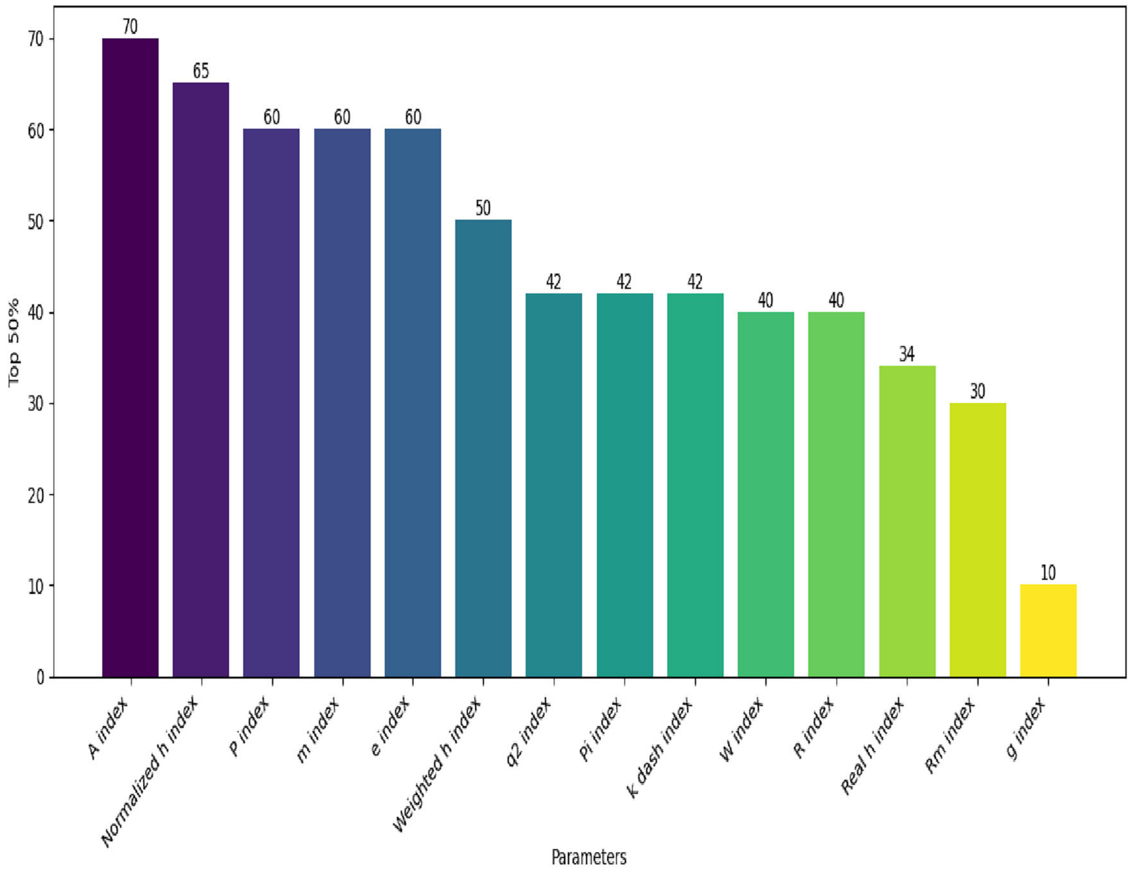

Fig. 7 Awardees percentage at top 50%

4.2 Summary of results for association among indices and awarding societies

As we delved further into analyzing the correlation between indices and awarding societies, we discovered intriguing patterns that illuminated nuanced relationships within the academic community. Importantly, our investigation went beyond a simple evaluation of each index in isolation. Instead, we aimed to unravel the distinct performance of each awarding society in retrieving award winners across different ranking thresholds. Within the complex web of these associations, the g index surfaced as intricately intertwined with the esteemed American Mathematical Society (AMS). This connection suggests a notable influence or recognition of the g index within the mathematical community, as substantiated by its formal affiliation with the esteemed American Mathematical Society (AMS).

Likewise, the R index demonstrated a clear association with the International Mathematical Union (IMU) society. This discovery not only underscores the global significance of the R index but also implies recognition by the IMU as a crucial metric within their sphere of influence. Shifting our focus to the q2 index and normalized h index, we uncovered a compelling alignment with the London Mathematical Society (LMS). This correlation suggests a resonance between the metrics encapsulated by these indices and the principles upheld by the LMS, indicating a potential preference or compatibility.

Remarkably, both the g index and e index revealed a mutual affiliation with the Norwegian Academy of Science and Letters (NASL). This dual connection implies a distinctive alignment of these indices with the values and standards upheld by the NASL, suggesting a specific acknowledgment or resonance within the Norwegian academic context. These intricate relationships underscore the significance of comprehending not just the performance of individual indices but also the contextual associations with awarding societies. Such insights contribute to a nuanced understanding of how various metrics are valued and recognized within specific academic and professional circles. Further exploration and analysis may unveil additional layers of understanding, enriching our insights into the dynamics between indices and awarding societies in the domain of mathematical recognition and achievement.

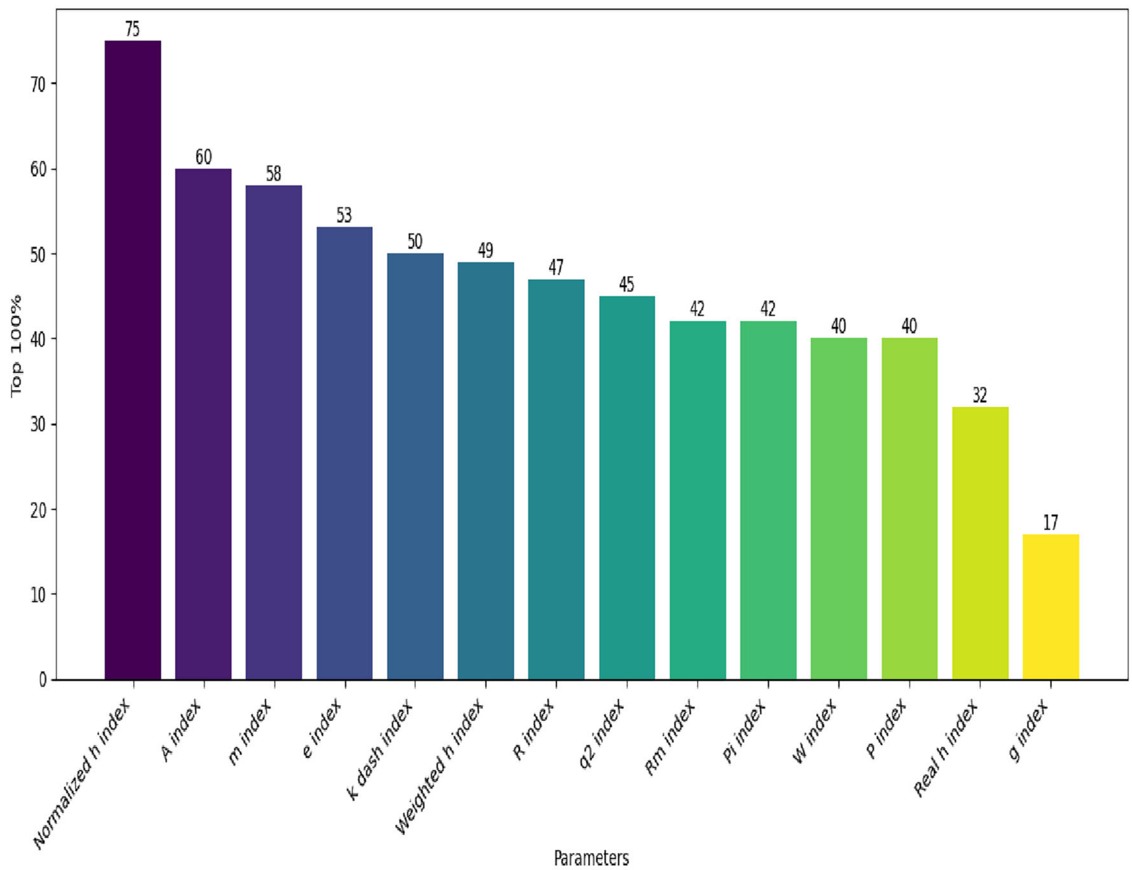

Fig. 8 Awardees percentage at top 100%

5 Conclusion

The current study conducts a systematic assessment of multiple scholarly indices, with particular emphasis on the A-index and normalized h-index, which have shown notable effectiveness in identifying high-impact researchers. Our analysis reveals that these indices can successfully retrieve over 60% of award-winning scholars, underscoring their capability to rank and recognize individuals who have made significant contributions to their fields. This performance demonstrates that certain h-index variations, especially the A-index and normalized h-index, address limitations inherent in the traditional h-index by offering a more nuanced and reliable measure of academic impact.

In addition, this research bridges an important gap in the literature by thoroughly examining the specific behaviors of these indices in realistic scenarios. Through evaluations across both award-winning and non-award-winning researchers, our findings shed light on the practical relevance of each metric in distinguishing influential scholars.

The findings of this research hold significant implications for both academics and decision-makers. The primary objective of ranking is to assess the scientific impact of researchers and consider them for tenure positions, post-doctoral and faculty positions. Decision-makers and human resource departments of any organization will utilize this information to make determinations regarding the recruitment of researchers, granting promotions, awarding prizes, inviting distinguished researchers as speakers, and allocating funds to individuals.

5.0.1 Limitations of study

Despite the abundance of parameters proposed for quantifying impact assessment in science, there is still a lack of universally accepted criteria or metrics. Although the results indicate that the A-index and normalized h-index are suitable for the field of mathematics, their effectiveness may not be consistent across other scientific disciplines. The aim of our study was not to promote a universal, one-size-fits-all evaluation approach. Rather, our objective was to examine the efficacy of specific metrics within a specific context: evaluation of award winners. In this particular context, we concentrated on a group of mathematicians as a case study to gauge the appropriateness of these metrics for this specific group. Consequently, the purview of our study is restricted to the assessment of awardees within the field of mathematics.

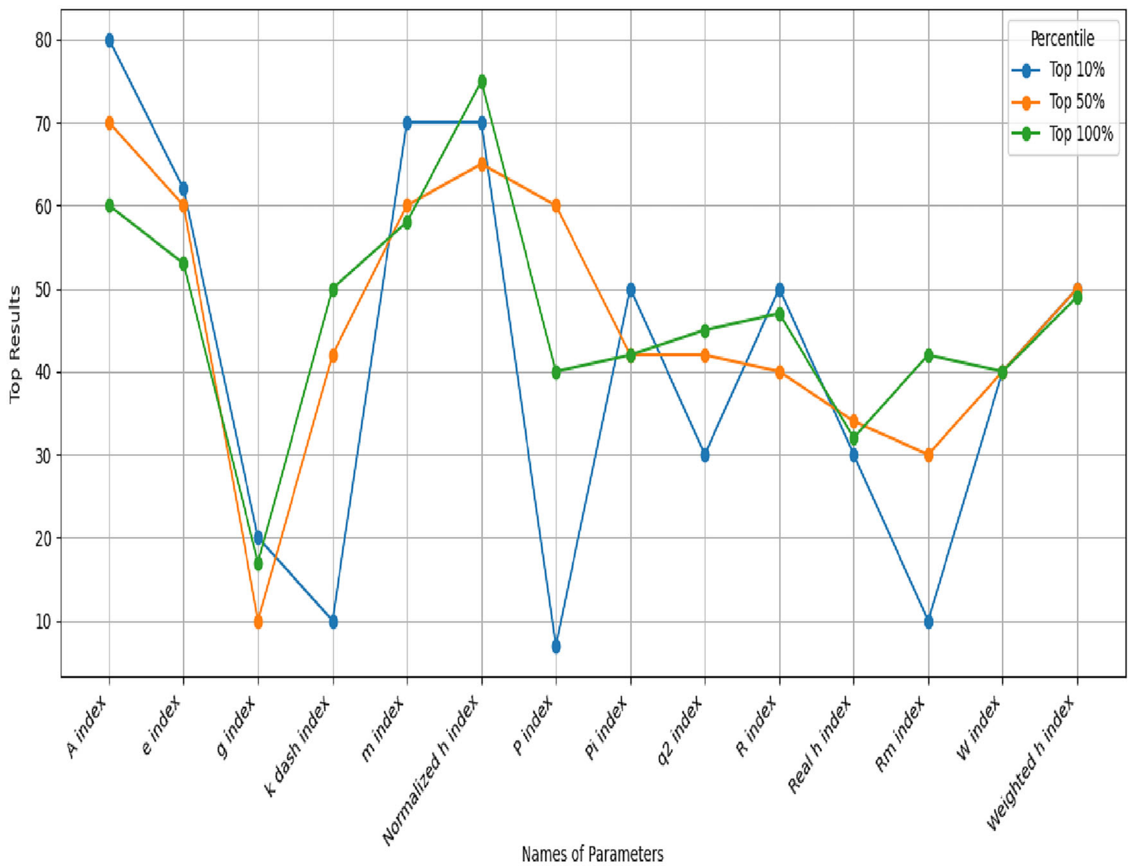

Fig. 9 Researchers score against each index

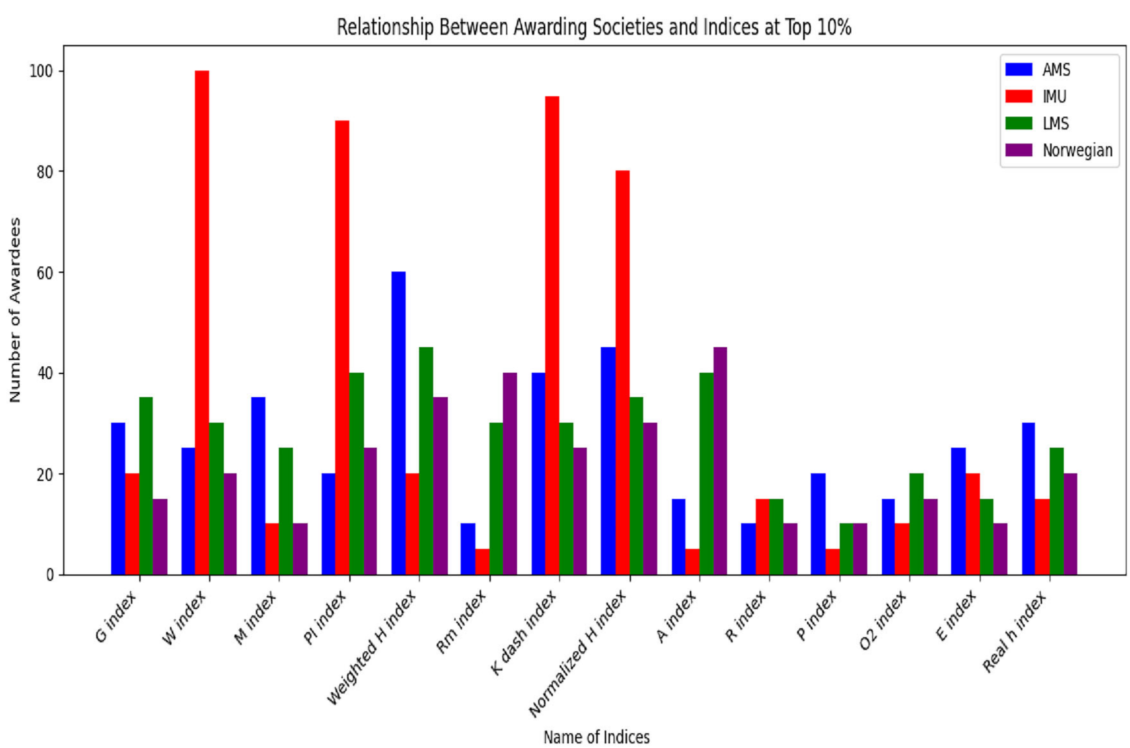

Fig. 10 Association between indices and awarding societies at top 10%

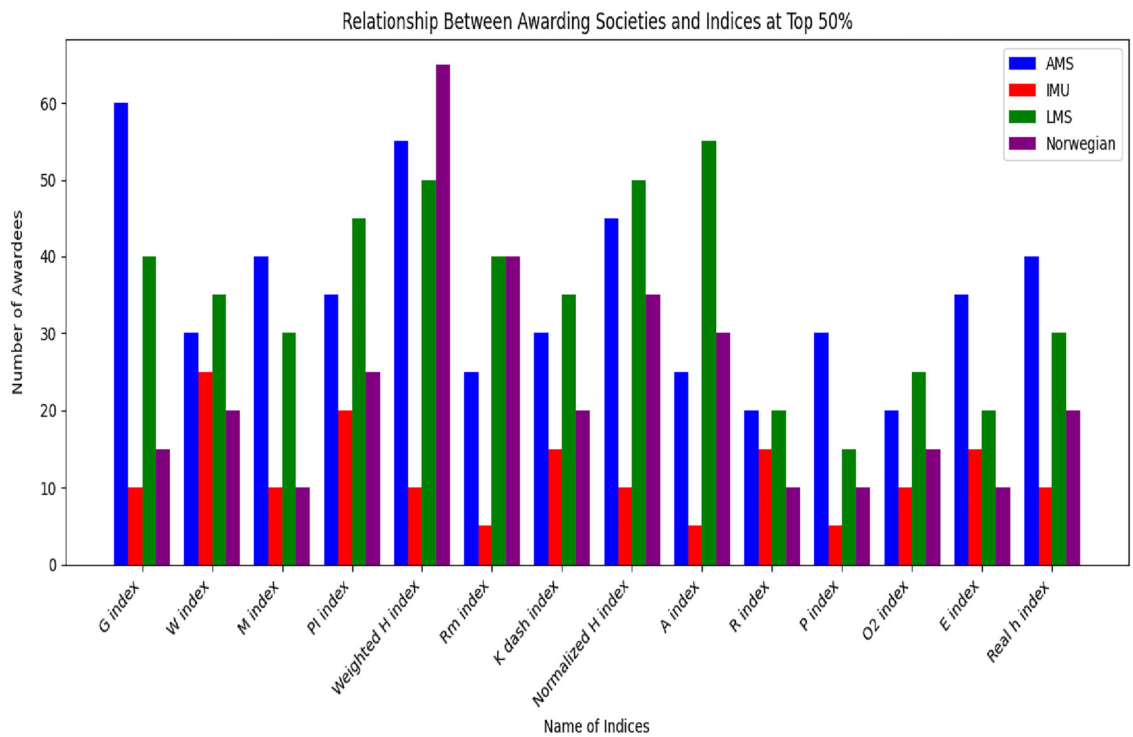

Fig. 11 Association between indices and awarding societies at top 50%

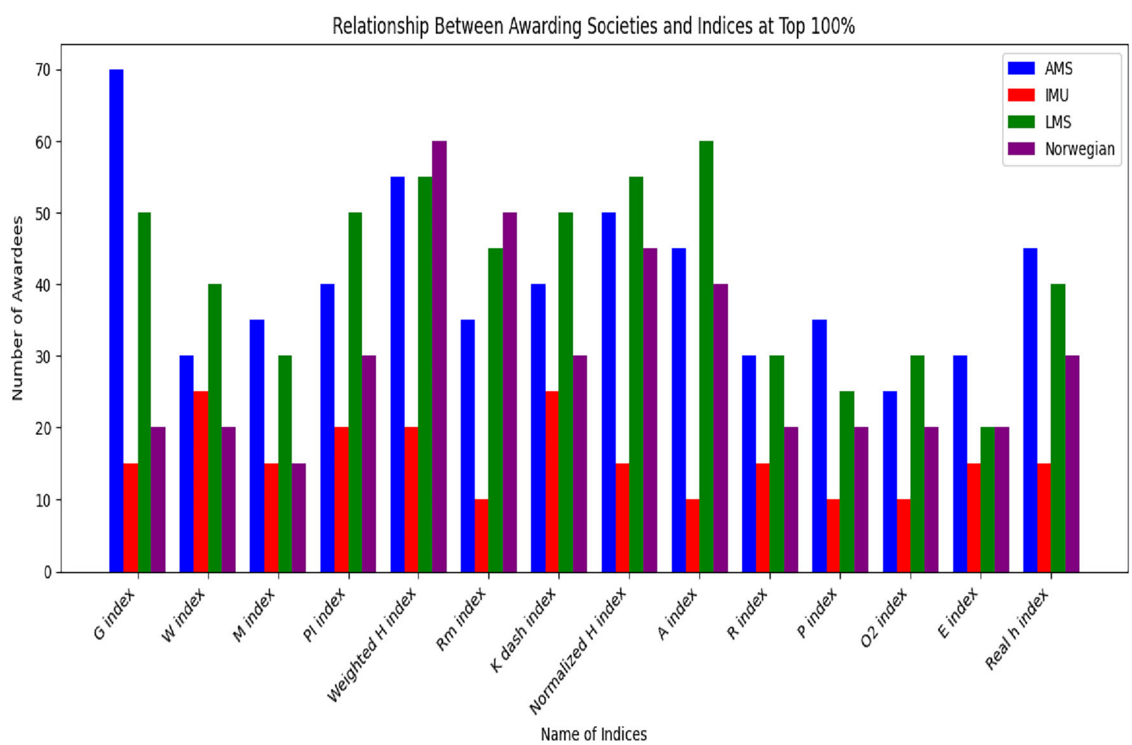

Fig. 12 Association between indices and awarding societies at top 100%

References

1. Mustafa, G., Usman, M., Yu, L., Afzal, M.T., Sulaiman, M., Shahid, A.: Multi-label classification of research articles using word2vec and identification of similarity threshold. Sci. Rep. **11**(1), 21900 (2021). https://doi.org/10.1038/s41598-021-01460-7
2. Mustafa, G., Usman, M., Afzal, M.T., Shahid, A., Koubaa, A.: A comprehensive evaluation of metadata-based features to classify research paper's topics. IEEE Access **9**, 133500–133509 (2021). https://doi.org/10.1109/ACCESS.2021.3115148
3. Mustafa, G., Rauf, A., Al-Shamayleh, A.S., Sulaiman, M., Alrawagfeh, W., Afzal, M.T., Akhunzada, A.: Optimizing document classification: unleashing the power of genetic algorithms. IEEE Access **4**, 1–1 (2023). https://doi.org/10.1109/ACCESS.2023.3292248
4. Mustafa, G., Rauf, A., Ahmed, B., Afzal, M.T., Akhunzada, A., Alharthi, S.Z.: Comprehensive evaluation of publication and citation metrics for quantifying scholarly influence. IEEE Access **11**, 65759–65774 (2023). https://doi.org/10.1109/ACCESS.2023.3290917
5. Vavryčuk, V.: Fair ranking of researchers and research teams. PLoS ONE **13**(4), 0195509 (2018). https://doi.org/10.1371/journal.pone.0195509
6. Usman, M., Mustafa, G., Afzal, M.T.: Ranking of author assessment parameters using logistic regression. Scientometrics **126**(1), 335–353 (2021). https://doi.org/10.1007/s11192-020-03769-y
7. Khokhlov, A.: How scientometrics became the most important science for researchers of all specialties. Mosc. Univ. Biol. Sci. Bull. **75**, 159–163 (2020). https://doi.org/10.3103/S0096392520040057
8. Ahmed, B., Li, W., Mustafa, G., Afzal, M.T., Alharthi, S.Z., Akhunzada, A.: Evaluating the effectiveness of author-count based metrics in measuring scientific contributions. IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3309416
9. Mustafa, G., Rauf, A., Al-Shamayleh, A.S., Ahmed, B., Alrawagfeh, W., Afzal, M.T., Akhunzada, A.: Exploring the significance of publication-age-based parameters for evaluating researcher impact. IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3304013
10. Ahmed, B., Wang, L., Al-Shamayleh, A.S., Afzal, M.T., Mustafa, G., Alrawagfeh, W., Akhunzada, A.: Machine learning approach for effective ranking of researcher assessment parameters. IEEE Access **11**, 133294–133312 (2023). https://doi.org/10.1109/ACCESS.2023.3336950
11. Wang, R., Lewis, M., Zheng-Pywell, R., Julson, J., Smithson, M., Chen, H.: Using the h-index as a factor in the promotion of surgical faculty. Heliyon **8**(4), e09319 (2022)
12. Mustafa, G., Rauf, A., Afzal, M.T.: Enhancing author assessment: an advanced modified recursive elimination technique (MRET) for ranking key parameters and conducting statistical analysis of top-ranked parameter. Int. J. Data Sci. Anal. **22**, 1–23 (2024). https://doi.org/10.1007/s41060-024-00545-6
13. Mustafa, G., Rauf, A., Al-Shamayleh, A.S., Afzal, M.T., Waqas, A., Akhunzada, A.: Defining quantitative rules for identifying influential researchers: insights from mathematics domain. Heliyon **10**(9), e30318 (2024)
14. Gogoglou, A., Sidiropoulos, A., Katsaros, D., Manolopoulos, Y.: The fractal dimension of a citation curve: quantifying an individual's scientific output using the geometry of the entire curve. Scientometrics **111**, 1751–1774 (2017). https://doi.org/10.1007/s11192-017-2285-2
15. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. **102**(46), 16569–16572 (2005). https://doi.org/10.1073/pnas.050765510
16. Ding, J., Liu, C., Kandonga, G.A.: Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. Scientometrics **122**, 1303–1322 (2020). https://doi.org/10.1007/s11192-020-03364-1
17. Bihari, A., Tripathi, S., Deepak, A.: A review on h-index and its alternative indices. J. Inf. Sci. **49**(3), 624–665 (2023). https://doi.org/10.1177/01655515211014478
18. Mokhtarpour, R., Khasseh, A.A.: Who is who in library and information science research? the integrative application of scholarly influence indicators. J. Librariansh. Inf. Sci. **52**(4), 1186–1196 (2020). https://doi.org/10.1177/0961000620907956
19. Lima, H., Silva, T.H., Moro, M.M., Santos, R.L., Meira Jr, W., Laender, A.H.: Aggregating productivity indices for ranking researchers across multiple areas. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 97–106 (2013)
20. Liang, R., Jiang, X.: Scientific ranking over heterogeneous academic hypernetwork. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 (2016). https://doi.org/10.1609/aaai.v30i1.10004
21. Lathabai, H.H., Prabhakaran, T.: Contextual ψ-index and its estimate for contextual productivity assessment. Scientometrics **128**(8), 4875–4886 (2023). https://doi.org/10.1007/s11192-023-04757-8
22. Sharma, K., Uddin, Z.: Measuring the continuous research impact of a researcher: the Kz index. arXiv preprint arXiv:2306.15677 (2023). https://doi.org/10.48550/arXiv.2306.15677
23. Basu, A., Banshal, S.K., Singhal, K., Singh, V.K.: Designing a composite index for research performance evaluation at the national or regional level: ranking central universities in india. Scientometrics **107**, 1171–1193 (2016). https://doi.org/10.1007/s11192-016-1935-0
24. Knight, J.: Academic impact rankings of neurosurgical units in the UK and Ireland, as assessed with the h-index. Br. J. Neurosurg. **29**(5), 637–643 (2015). https://doi.org/10.3109/02688697.2015.1039488
25. Rubem, A.P.D.S., Moura, A.L., Mello, J.C.C.B.: Comparative analysis of some individual bibliometric indices when applied to groups of researchers. Scientometrics **102**, 1019–1035 (2015). https://doi.org/10.1007/s11192-014-1428-y
26. Schreiber, M.: An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the a-index, and the r-index. J. Am. Soc. Inform. Sci. Technol. **59**(9), 1513–1522 (2008). https://doi.org/10.1002/asi.20856
27. Bi, H.H.: Four problems of the h-index for assessing the research productivity and impact of individual authors. Scientometrics **128**(5), 2677–2691 (2023). https://doi.org/10.1007/s11192-022-04323-8
28. Koltun, V., Hafner, D.: The h-index is no longer an effective correlate of scientific reputation. PLoS ONE **16**(6), 0253397 (2021). https://doi.org/10.1371/journal.pone.0253397
29. Ayaz, S., Masood, N.: Comparison of researchers' impact indices. PLoS ONE **15**(5), 0233765 (2020)
30. Ameer, M., Afzal, M.T.: Evaluation of h-index and its qualitative and quantitative variants in neuroscience. Scientometrics **121**(2), 653–673 (2019). https://doi.org/10.1007/s11192-019-03209-6
31. Raheel, M., Ayaz, S., Afzal, M.T.: Evaluation of h-index, its variants and extensions based on publication age & citation intensity in civil engineering. Scientometrics **114**, 1107–1127 (2018). https://doi.org/10.1007/s11192-017-2633-2
32. Mukherjee, B.: Ranking Indian universities through research and professional practices of national institutional ranking framework (NIRF): a case study of selected central universities in India. J Indian Library **52**(4), 454 (2019)
33. Wilkes, F.A., Akram, H., Hyam, J.A., Kitchen, N.D., Hariz, M.I., Zrinzo, L.: Publication productivity of neurosurgeons in great Britain and Ireland. J. Neurosurg. **122**(4), 948–954 (2015). https://doi.org/10.3171/2014.11.JNS14856

34. Sharma, B., Boet, S., Grantcharov, T., Shin, E., Barrowman, N.J., Bould, M.D.: The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: a province-wide study. Surgery **153**(4), 493–501 (2013). https://doi.org/10.1016/j.surg.2012.09.006
35. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., Delgado López-Cózar, E.: Google scholar, microsoft academic, scopus, dimensions, web of science, and opencitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics **126**(1), 871–906 (2021). https://doi.org/10.1007/s11192-020-03690-4
36. Jin, B., Liang, L., Rousseau, R., Egghe, L.: The R-and AR-indices: complementing the h-index. Chin. Sci. Bull. **52**(6), 855–863 (2007). https://doi.org/10.1007/s11434-007-0145-9
37. Zhang, C.-T.: The e-index, complementing the h-index for excess citations. PLoS ONE **4**(5), 5429 (2009). https://doi.org/10.1371/journal.pone.0005429
38. Egghe, L., Rousseau, R.: An h-index weighted by citation impact. Inf. Process. Manage. **44**(2), 770–780 (2008). https://doi.org/10.1016/j.ipm.2007.05.003
39. Chen, D.-Z., Huang, M.-H., Fred, Y.Y.: A probe into dynamic measures for h-core and h-tail. J. Informet. **7**(1), 129–137 (2013). https://doi.org/10.1016/j.joi.2012.10.002
40. Bornmann, L., Mutz, R., Daniel, H.-D.: Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. J. Am. Soc. Inform. Sci. Technol. **59**(5), 830–837 (2008). https://doi.org/10.1002/asi.20806
41. Sidiropoulos, A., Katsaros, D., Manolopoulos, Y.: Generalized hirsch h-index for disclosing latent facts in citation networks. Scientometrics **72**(2), 253–280 (2007). https://doi.org/10.1007/s11192-007-1722-z
42. Aziz, N.A., Rozing, M.P.: Profit (p)-index: the degree to which authors profit from co-authors. PLoS ONE **8**(4), 59814 (2013). https://doi.org/10.1371/journal.pone.0059814
43. Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., Herrera, F.: q2-index: quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. J. Informet. **4**(1), 23–28 (2010). https://doi.org/10.1016/j.joi.2009.06.005
44. Panaretos, J., Malesios, C.: Assessing scientific research performance and impact with single indices. Scientometrics **81**, 635–670 (2009). https://doi.org/10.1007/s11192-008-2174-9
45. Guns, R., Rousseau, R.: Real and rational variants of the h-index and the g-index. J. Informet. **3**(1), 64–71 (2009). https://doi.org/10.1016/j.joi.2008.11.004
46. Wu, Q.: The w-index: a measure to assess scientific impact by focusing on widely cited papers. J. Am. Soc. Inform. Sci. Technol. **61**(3), 609–614 (2010). https://doi.org/10.1002/asi.21276

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.