publications

MDPI

Article

Measuring Group Performance Fairly: The h-Group, Homogeneity, and the α-Index

Roberto da Silva [1,*], José Palazzo M. de Oliveira [2,3] and Viviane Moreira [2,3]

[1] Instituto de Física, Universidade Federal do Rio Grande do Sul (UFRGS), P.O. Box 15051, Porto Alegre 91501-970, RS, Brazil
[2] Instituto de Informática, Universidade Federal do Rio Grande do Sul (UFRGS), P.O. Box 15064, Porto Alegre 91501-970, RS, Brazil; palazzo@inf.ufrgs.br (J.P.M.d.O.); viviane@inf.ufrgs.br (V.M.)
[3] Postgraduate Program in Computing (PPGC), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, RS, Brazil
* Correspondence: rdasilva@if.ufrgs.br

Abstract

Ranking research groups plays a crucial role in various contexts, such as ensuring the fair allocation of research grants, assigning projects, and evaluating journal editorial boards. In this paper, we analyze the distribution of h-indexes within research groups and propose a single metric to quantify their overall performance, termed the *α-index*. This index integrates two complementary aspects: the homogeneity of members' *h-indexes*, captured by the Gini coefficient (g), and the *h-group*, an extension of the individual *h-index* to groups. By combining both uniformity and collective research output, the *α-index* provides a consistent and equitable metric for comparative evaluation, essentially calculated as the average relative *h-group* multiplied by $(1 - g)$ and normalized by the maximum value of this quantity across all analyzed groups. We describe the full procedure for computing the index and its components and illustrate its application to computer science conferences, where program committees are compared through a resampling procedure that ensures fair comparisons across groups of different sizes. Additional results are presented for postgraduate programs, further demonstrating the method's applicability. Correlation analyses are used to establish rankings; however, our primary goal is to recommend a fairer index that reduces deviations from those currently used by governmental agencies to evaluate conferences and graduate programs. The proposed approach offers a more nuanced assessment than simply averaging members' *h-indexes* and can be applied broadly–for example, to university departments and research councils–contributing to a more equitable distribution of research funding, an issue of increasing importance.

Keywords: Gini coefficient; h-index; metrics in science; bibliometrics

check for updates

Academic Editor: Andrew Kirby

Received: 13 September 2025
Revised: 2 December 2025
Accepted: 3 December 2025
Published: 11 December 2025

Citation: da Silva, R., de Oliveira, J. P. M., & Moreira, V. (2025). Measuring Group Performance Fairly: The h-Group, Homogeneity, and the α-Index. *Publications*, *13*(4), 66. https://doi.org/10.3390/publications13040066

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Universal laws for citations have been among the most discussed topics in the academic community over the past decades (Hirsch, 2005; Laherrère & Sornette, 1998; Redner, 1998). Rankings, in turn, constitute another fundamental concept for the academic community, as they permeate several processes, ranging from university rankings to, for example, the admission of students to universities da Silva et al. (2016), and, in a bibliometric context, to the classification of researchers Batista et al. (2006). Bibliometric rankings can also be useful for allocating research grants in a fair manner and for assessing journals according to the quality of their editorial boards. In the scope of this paper, the term *quality* refers to

research output as measured by scientific publications. Thus, how authors' citations can be captured to build such bibliometric rankings is a problem that deserves careful attention.

Numerical data on the distribution of citations has been extensively explored by the scientific community, and universal laws have been established. Based on the ISI (Institute for Scientific Information) database, Laherrère and Sornette (1998) suggest that the number of papers with x citations decays as a power law $N(x) \sim x^{-\alpha}$ where $\alpha \sim 3$. Similarly, Redner (1998) found a stretched exponential form $N(x) \sim \exp[-(x/x_0)^{\beta}]$ with $\beta \sim 0.3$ when analysing data from the 1120 most-cited physicists between 1981 and 1997. On the other hand, Tsallis and de Albuquerque (2000) showed that citation distributions are well described by curves derived from nonextensive formalism.

Hirsch (2005) reduced the complexity of the data distribution to quantify the importance of a scientist's research output into a single measure known as the *h-index*. Despite being controversial, the *h-index* is widely employed by many research funding agencies and universities all over the world. Hirsh's simple idea is that a publication is good as long as it is cited by other authors; i.e., "a scientist has index h if h of their N_p papers has at least h citations each. The other $(N_p - h)$ papers have $\leq h$ citations each, with $0 \leq h \leq N_p$."

There are alternatives to the use of the *h-index*; Braun (2004), for example, uses the total number of citations to quantify research performance. However, in this paper, we have opted to use the *h-index* because it is less prone to being inflated by a small number of big hits or by the eminence of co-authors. Since its proposal, it has become widely accepted and has been employed as the basis for many scientometrics and bibliometrics research. Moreover, since it is a flexible index, authors explore its range. For example, the authors of Batista et al. (2006) show that accounting for the number of coauthors in the h-core papers used to compute a researcher's h-index makes it possible to compare researchers across different scientific fields.

In recent decades, bibliometrics has evolved substantially in the domains of *equity and diversity indicators*, *modern normalization methods*, *network-based approaches*, and *epistemological reflections* on research evaluation.

Concerning *equity and diversity*, the discussion has been greatly shaped by the *Leiden Manifesto*, which proposes ten principles to guide the responsible use of research metrics and to promote fairness, transparency, and contextualization in evaluation processes (Hicks et al., 2015). These principles are essential to mitigate systemic biases and to avoid reinforcing structural inequalities in scientific assessment.

Regarding *modern citation normalization*, Waltman and van Eck provided both a conceptual classification and detailed empirical analyses. Their work on source-normalized indicators demonstrates that these methods often outperform traditional field-normalized metrics, as they are better suited to handle the diverse citation practices across scientific fields Waltman and van Eck (2013a, 2013b). These contributions remain among the most influential in contemporary citation normalization research.

In the area of *network- and collaboration-based approaches*, Newman's studies on scientific collaboration networks revealed that co-authorship structures display small-world properties, strong clustering, and characteristic patterns of connectivity (Newman, 2001). Extended analyses further explored temporal evolution, centrality measures, and distances among researchers, providing a deeper understanding of how scientific communities are organized (Newman, 2004). Other authors (Silva et al., 2013) have used network-based bibliometric analyses to quantify the interdisciplinarity of scientific journals and research fields.

From an *epistemological perspective*, Biagioli and Lippman's edited volume *Gaming the Metrics* investigates how metric-based evaluation systems can generate perverse incentives, leading to misconduct, citation rings, and other forms of strategic manipulation (Biagioli

& Lippman, 2020). That work highlights the non-neutral nature of metrics and their interaction with institutional pressures. Similarly, Moed and Halevi argue for the need for a multidimensional assessment framework, showing that no single metric can adequately capture the complexity of scientific performance and that different indicators are relevant for different evaluation contexts (Moed & Halevi, 2015).

Together, these contributions illustrate the richness and complexity of contemporary bibliometric scholarship, emphasizing that responsible research evaluation requires a combination of quantitative indicators, qualitative judgment, epistemic awareness, and sensitivity to equity.

In this context, the importance of analyzing groups of researchers is evident, as such evaluations are directly related to the distribution of research grants, the composition of editorial boards, the assessment of postgraduate programs, and other institutional processes. Among recent works, we can mention the study of university chemistry groups in the Netherlands, which shows how group performance correlates with journal impact (van Raan, 2012). In their study, Thaule and Strand (2018) compare normalization methods and bibliometric evaluation approaches. Rosas et al. (2011) apply bibliometric methods to clinical trial networks funded by a U.S. agency, assessing their research presence, performance, and impact. This study addresses the evaluation of large-scale research programs, which can be particularly useful if the analyzed "group" represents a broader program or consortium. Biagetti (2022) critically discusses the use of bibliometric indicators and advocate for their complementarity with peer review and other qualitative methods.

The problem addressed in the present work is how to characterize and classify a group of researchers by quantitatively analyzing the *h-indexes* of its individual members. The method we propose assumes that quality cannot be characterized just by a high average *h-index* for the group, but also by its homogeneity. Our rationale is that a group can have a high average *h-index* just by having one very productive researcher. However, a homogeneous group with an equivalent *h-index* will be better, as homogeneity denotes greater robustness of the group.

We introduce a method to measure the scientific research output of a group of researchers. The proposed method quantifies the quality of a group using a parameter that we call the *α-index*. The *α-index* of a group is based on two concepts:

- The *h-group*, which is an extension of the *h-index* for groups. It is measured by taking the maximum number of researchers in the group, satisfying $h\text{-}index \geq h\text{-}group$. The remaining researchers in the group have $h\text{-}index < h\text{-}group$.
- A known statistic employed to demonstrate the social inequality of a country, the Gini coefficient (Gini, 1921; Lopes et al., 2011).

An important aspect in the fields of computer science and engineering is that not only journal publications are valued; qualified conferences and workshops also play a crucial role Zhuang et al. (2007). Consequently, assessing the quality of conferences is essential for an adequate evaluation of a researcher's output. In this work, we show how the proposed *α-index* can be applied, first, to evaluate the quality of a conference based on the *h-indexes* of the members of its Technical Program Committee (TPC), and, second, to examine a completely different context: Brazilian postgraduate programs (BPPs) in Physics–also through the *h-index* of their members, whose scientific production is predominantly based on journal publications.

Our method was designed to enable fair comparisons between groups of different sizes, making it suitable for analyzing both TPCs and BPPs. The results obtained with the proposed *α–index* are consistent: the relative ordering of the groups remains stable even when considering subsets or supersets of the full set of groups.

In our analysis, we gathered and organized bibliometric data for the seven conferences held in 2007 and for the nine postgraduate programs evaluated in 2010. This dataset includes the individual *h-index* of each TPC and BPP member, as well as the citation counts of their publications.

The remainder of this paper is organized as follows. Section 2 presents a pedagogical study of bibliometric data using the information collected from the seven conferences. We discuss several stylized facts regarding citation and h-index distributions, in order to verify whether the dataset reproduces the universal bibliometric patterns reported in the literature. We also present the classification of these conferences according to CAPES (http://www.capes.gov.br, accessed on 28 September 2007), the Brazilian agency responsible for evaluating the quality of graduate programs (Almeida Guimarães & Chaves Edler de Almeida, 2012; Lima, 2025). Section 3 shows that the Gini coefficient is a natural measure of the homogeneity of the program committees of scientific conferences, by analogy with the homogeneity of wealth distribution in a population. We also define the concept of the *h-group*. These two notions allow us to define the index that will be explored in our results: the group's α*-index*.

It is important to note that Sections 2 and 3 serve as preparatory material. For simplicity, these sections use only the TPC dataset to illustrate the concepts and statistical methods introduced. After this groundwork is established, we turn to our main quantity, the α-index, and analyze it in detail using both datasets. The results for TPCs and BPPs are then presented separately in Sections 4 and 5, respectively.

First, we analyze the TPCs in computer science. We compute the α-index for each conference and perform ranking tests based on the CAPES classification. Then, to complement this analysis—and to show that the index can be applied consistently across different contexts—we present a second application that serves as a validation of the approach, calculating the α-index for the BPPs in Physics. In this case, we briefly describe the dataset, compute the α-index, and show that it also aligns well with the CAPES ranking (accessed at http://www.capes.gov.br on 17 June 2010), similarly to what was observed for the TPCs. Finally, Section 6 presents the conclusions and outlines possible extensions of the model.

The results indicate that the methodology is effective for both datasets, showing a strong correlation with the CAPES classification.

2. Preliminaries, Descriptive Statistics, and Stylized Bibliometric Facts: The Case of Computer Science Conferences

The data used for the TPCs in this study refer to an earlier period (2007), and the corresponding analyses were refined continuously up to the present submission. Although the dataset is not up to date, this choice ensures that it is unaffected by distortions arising from the COVID-19 pandemic or other external factors. The resulting h-index values thus reflect the conditions of that period and should not be interpreted as representative of the currently inflated citation rates or the recent surge in scientific output. For ethical reasons, the names of universities, conferences, locations, and individual researchers have been omitted to preserve the privacy and integrity of all parties involved.

CAPES has defined a system for classifying the estimated quality of publication venues. The system is called *Qualis*, and it grades venues into three categories: A, B, or C. According to this grading scheme, A is the highest quality, and it is usually assigned to top international conferences. The criteria analysed include the number of editions of the conference and its acceptance rate.

Table 1 presents the data collected for seven conferences, all corresponding to the same validity period as the official ranking. The h-indexes were obtained using the free software *Publish or Perish* (http://www.harzing.com/resources.htm, accessed on 28

September 2007), which extracts citation data from the Google Scholar service (http://scholar.google.com, also accessed on 28 September 2007).

Table 1. Conferences used for this work and their classification. The second column shows the average *h-index* for the TPC with the associated standard error $(var(h)/n)^{1/2}$.

Conference	$\langle h \rangle \pm (var(h)/n)^{1/2}$	**#TPC Members**
Conf. A (A)	12.78 ± 0.65	207
Conf. B (C)	11.92 ± 1.60	27
Conf. C (B)	11.63 ± 1.55	67
Conf. D (A)	10.10 ± 0.66	102
Conf. E (B)	08.07 ± 0.83	87
Conf. F (A)	07.94 ± 0.69	39
Conf. G (C)	07.56 ± 2.39	16

As a starting point, we explore some preliminary statistics about these conferences. It is interesting to check similarities between the properties obtained from the TPC population and the properties expected from the general scientific population. The first analysis was to plot the number of citations to papers written by the TPC members as a function of their *h-indexes*. These plots are shown in Figure 1.

In all cases presented in Figure 1, we found that the number of citations the authors have has a quadratic dependence on their h-indexes, as obtained in scientific databases such as ISI (see, for example, Laherrère and Sornette (1998); Redner (1998)). In order to measure the α exponent, we separated our data according to the classification of the conference (A, B or C) assigned by CAPES. For each set of conferences, we analysed the expected relation $x \sim h^{\alpha}$, where x is the number of citations of the author and h is the corresponding *h-index*. In a log–log plot, shown in Figure 1, we measured the slope. The results were $\alpha = 2.08(3)$, 2.12(3), and 2.15(6) for conferences A, B, and C, respectively. This result corroborates Hirsh's theory (Hirsch, 2005), in which $\alpha = 2$.

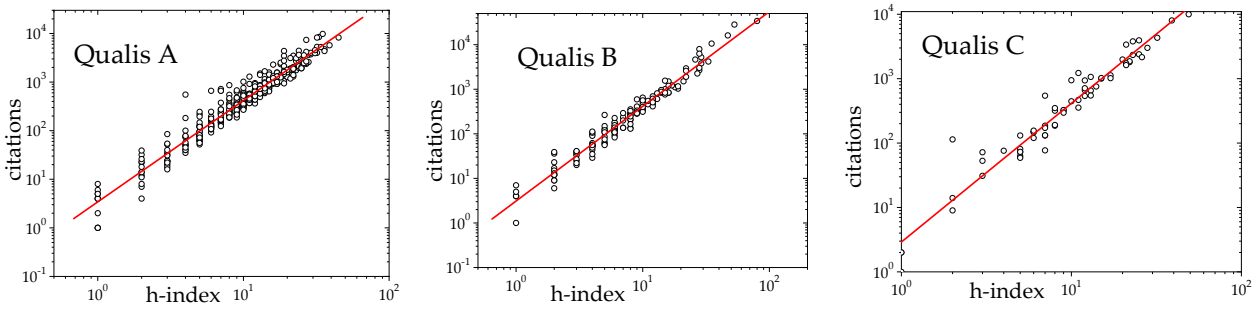


Figure 1. Citations versus *h-index* for conferences Qualis A, B, and C.

We analyzed the distribution of citations in order to compare the features of our data against the properties found in other scientific populations. The analysis takes all TPC members into consideration (combining A, B, and C conferences). The idea was to verify whether the distribution of the number of citations, denoted by x, for TPC members of computer science conferences follows a stretched exponential form:

$$N(x) \propto \exp\left[-(x/x_0)^{\beta}\right] \tag{1}$$

as claimed by Laherrere and Sornette in (Laherrère & Sornette, 1998). In their study, they found $\beta \sim 0.3$, which can be determined by plotting a histogram with the number of citations (x), as shown in Figure 2 (left plot).

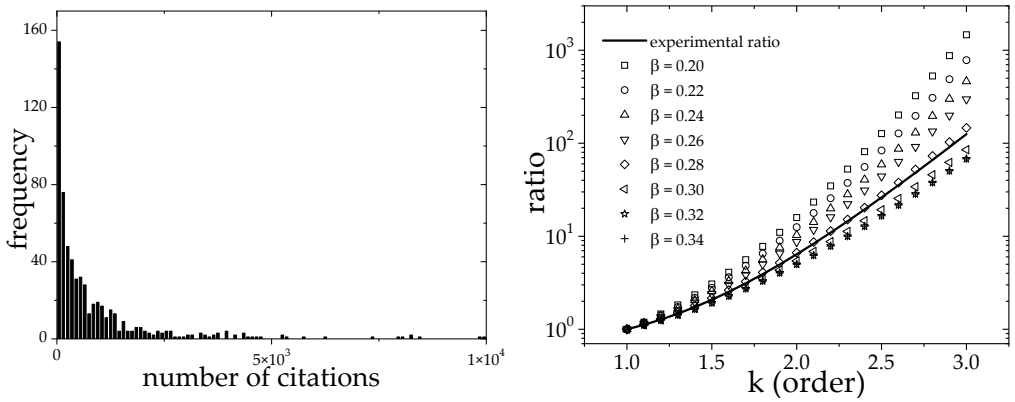


Figure 2. (**Left plot**) Experimental distribution of the number of citations for data obtained from conferences Qualis A, B, and C. (**Right plot**) A comparison between the exact moments (Equation (2)) obtained for different β values and the real moment, i.e., obtained from the collected data (see Equation (3)).

Estimating β directly from the data in Figure 2 (left plot) can be achieved by determining the slope of the linear fit of $\log(\log(N(x)))$ versus log x. Nevertheless, this procedure may yield imprecise results. Therefore, we propose looking at the exact ratio $M_k = \langle x^k \rangle / \langle x \rangle^k$, where $\langle x^k \rangle$ are the moments of the distribution given by Equation (1). For example, when $k = 1$, $\langle x \rangle$ corresponds to the average of the distribution given by Equation (1). The solution we adopted was to vary β so as to find the best approximation to M_k in relation to the experimental ratios R_k calculated by Equation (3). Thus, the values for M_k can be analytically calculated and do not depend on the parameter x_0:

$$M_k = \frac{\Gamma\left(\frac{k+1}{\beta}\right)\Gamma\left(\frac{1}{\beta}\right)^{k-1}}{\Gamma\left(\frac{2}{\beta}\right)^{k}}, \tag{2}$$

where $\Gamma(z)$ is the gamma function $\Gamma(z) = \int_0^\infty t^{z-1}e^t dt$.

We then calculate M_k for values of k between 1 and 3, using a lag of $\Delta k = 0.1$, and different values of $\beta = 0.20, 0.22, \ldots, 0.34$ (see right plot in Figure 2) were considered in the search for a best match to the experimental ratio R_k given by Equation (3).

$$R_k = \frac{(1/n)\sum_{i=1}^{n} x_i^k}{[(1/n)\sum_{i=1}^{n} x_i]^k}, \tag{3}$$

where n denotes the number of TPC members and is represented by a continuous curve in the same plot. We can observe that the best match is found when $\beta = 0.28$, corroborating the expected behavior as described in Laherrère and Sornette (1998).

This brief analysis shows that the statistical properties related to the distribution of the number of citations and its relationship with the *h-index* are similar to what was observed in other scientific societies.

Let us also analyze some aspects related to the *h-index* distributions from TPC members of computer science conferences. A histogram of the *h-index* for all collected conferences is illustrated in Figure 3. Many empirical fits were tested (log-normal, gamma, and other non-symmetric functions). Because of the characteristics of the data, a normal fit was not attempted. An excellent fit was found by using a function that comes from the Chromatography literature, known as Giddings distribution (Giddings & Eyring, 1954), defined in the following equation:

$$H(h) = H_0 + \frac{A}{w}\sqrt{\frac{h_c}{h}}\, I_1\left(\frac{2\sqrt{h_c h}}{w}\right) \exp\left(\frac{-h - h_c}{w}\right), \tag{4}$$

where $I_1(x)$ is the modified Bessel function, which is described in the integral form by $I_1(x) = \frac{1}{\pi}\int_0^{\pi} e^{x\cos\theta}\cos\theta d\theta$. Apart from the difficulty in analytically evaluating this function, the fit is numerical and easy to be performed. The fitted values were $H_0 = 0$, $h_c = 10.44$, $w = 2.518$, and $A = 0.91$. The function given by Equation (4) is the distribution in t, representing the chance that one solute molecule will be eluted from the bottom of the column in a phenomenon of passage of substances through a chromatographic column (see Giddings and Eyring (1954) for further details). However, fitting a distribution with four parameters is not simple. A more intuitive formula is to consider the citation distribution given by Equation (1). By considering Hirsh's relation $x = ah^2$, using (1) followed by normalization, we obtain an *h-index* distribution as given by da Silva et al. (2012):

$$H_{new}(h) = \frac{2a\beta h}{x_0\Gamma(1/\beta)}\exp\left[-\left(\frac{ah^2}{x_0}\right)^{\beta}\right] \tag{5}$$

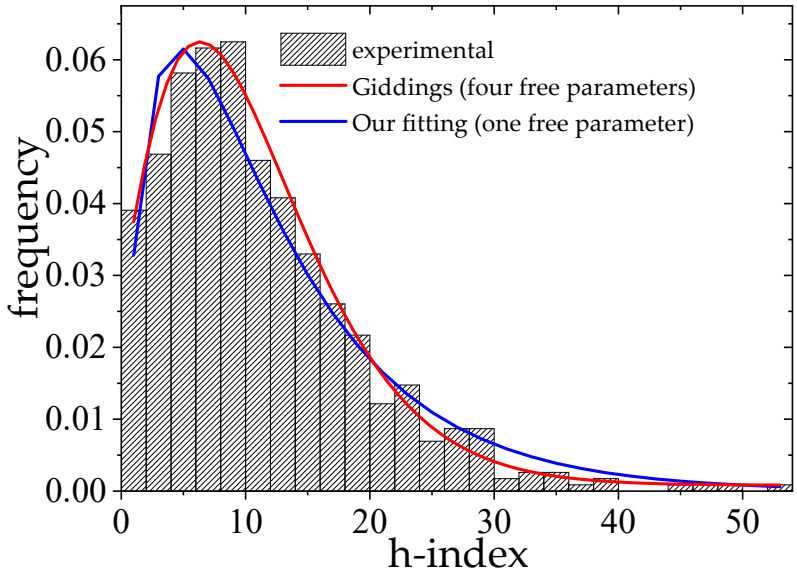


Figure 3. Histogram of the *h-index* of TPC members of all conferences. The best fit (red curve) is the Giddings function (four free parameters). In blue, we can see our proposed fit with just one parameter.

Thus, by considering a simple linear fit in a plot of x as function h^2 we obtain the slope a that for our conferences is $a = 5.71$ (recalling the Hirsch relation, $x = ah^2$). Since we also have β previously calculated, varying x_0 (from $x_{min} = 1$ to $x_{max} = 90$) we find the best fit, denoted by the blue curve in Figure 3. The best value found for x_0 by minimizing the least square function was $x_0 = 51$. Γ is the same gamma function already described in Equation (2).

The results of this analysis show a non-symmetric distribution of the *h-index* in the program committee of the conferences. However, is this indeed a good feature? In fact, we expect a good conference to have a homogeneous committee composed of young promising researchers with good *h-indexes* and also experienced researchers with a good *h-index* achieved through a sound scientific career. We do not consider a TPC composed of a few leading scientists padded up with lower-qualified researchers good. Thus, in a second investigation, we analyze the *h-index* distribution for each conference. First, it would be interesting to know if any of the conferences present a normal distribution of the h-indexes of their TPC members. Using a traditional Shapiro–Wilk (SW) normality test (see results in Table 2), we tested the normality level of each conference studied. The conferences Conf. F and Conf. C (the latter is normal at a much lower level) were considered to be normally distributed, at a level of 5%.

Table 2. Normality analysis of TPC members' *h-indexes*.

Conference	Normality	*p*-Value	Kurtosis	Skewness
Conf. D (A)	non-normal	0.00092	0.47117	0.70534
Conf. A (A)	non-normal	0.00000	9.57315	1.99863
Conf. F (A)	Normal	0.45027	−0.11212	0.40324
Conf. E (B)	non-normal	0.00000	2.63593	1.58998
Conf. C (B)	non-normal	0.00000	13.74526	3.14864
Conf. B (C)	Normal	0.06612	−0.24953	0.63032
Conf. G (C)	non-normal	0.00007	7.99799	2.15824

Figure 4 shows a graphical comparison between the *h-index* histogram for Conf. F, which is remarkably normal, and for Conf. A, which is remarkably non-normal.

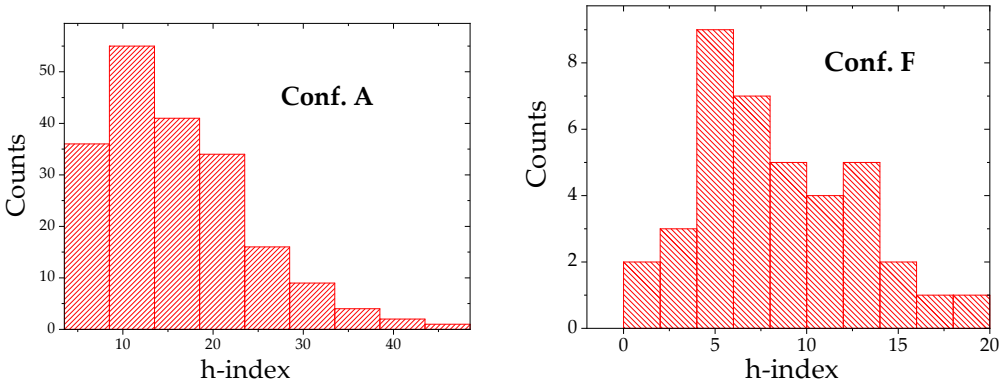


Figure 4. *h-index* histograms. The **left plot** shows a non-normal conference and the **right plot** shows a conference that is normal according to the SW test.

Table 2 also shows some other statistics. The third column presents the *p-values* (the greatest value to be attributed to type 1 error for which the normality test would not be rejected). The fourth column shows the kurtosis, which is calculated as in Equation (6), and measures the weight of the tail of the distribution. The fifth column shows the skewness, which is calculated as in Equation (7), and measures the symmetry of the distribution.

$$\text{kurtosis} = \frac{\langle (h - \langle h \rangle)^4 \rangle}{\langle (h - \langle h \rangle)^2 \rangle^2} - 3, \tag{6}$$

$$\text{skewness} = \frac{\langle (h - \langle h \rangle)^3 \rangle}{\langle (h - \langle h \rangle)^2 \rangle^{3/2}}. \tag{7}$$

For a Gaussian distribution, we expect kurtosis and skewness to approach zero. We can observe that Conf. A is a conference that does not follow a normal distribution, characterized by a high *p*-value (0.0000), which is more expected of a natural *h-index* distribution. It presents a heavy tail (kurtosis = 9.57315) and a good asymmetry (skewness = 1.99863) in relation to its mean value. As we report in Section 4, Conf. A is the best among the conferences analyzed with our proposed *index*.

There is no conclusive indication regarding the relationship between the normality of the TPC *h-index* distribution and the quality of the conference. However, according to our method (presented in the Section 3) the best conferences are predominantly non-normal. This suggests that good conferences will likely have many researchers with high *h-indexes* (a characteristic of heavy-tailed distributions), despite many of the *h-indexes* being concentrated around a mean value. Normality is an interesting aspect, but it is not the most adequate to determine quality. A good metric must take into consideration the homogeneity

of a group, but, at the same time, it cannot lose focus of the magnitude of *h-indexes* of the members. It is also important to allow comparisons among groups of different sizes.

Homogeneity, together with a reasonable *h*-index definition for groups (see Egghe (2008); Schubert (2007)), is the main requirement for a good research group, such as a TPC, a postgraduate program, or even the editorial board of a journal. These aspects are explored in greater detail in the next sections.

3. The Gini Coefficient and the *h-Index* of a Group

The notion of a high-quality group, in any context, presupposes the excellence of its individual members. In certain cases, however, it is not sufficient for a group to include merely a few highly productive individuals alongside others with limited academic output. Homogeneity—understood as a consistent level of academic productivity among members—is also a desirable attribute. This is particularly relevant for research groups such as program committees (TPCs) and editorial boards, where a more homogenous level of expertise contributes to the fairness and consistency of paper evaluations.

We suggest that homogeneity enhances the credibility and effectiveness of such evaluative bodies. Admission to a journal's editorial board or a conference's TPC should therefore be contingent upon the researcher having attained a scholarly profile commensurate with the quality standards of the venue. Given that publication venues vary in their academic rigor, participation as a reviewer or committee member should not be assumed as a default entitlement, but rather earned through demonstrated academic achievement appropriate to the venue's expectations. An interesting statistic to measure the equality of members in a group comes from the Social Economics literature, the Gini coefficient (Gini, 1921; Lopes et al., 2011).

In its original formulation, the Gini coefficient (which is a number in the interval [0, 1]) was designed to quantify inequalities in the distribution of wealth within a country. The lower the Gini coefficient, the more equal the wealth distribution. The highest known Gini coefficient is Namibia's (0.707) while the lowest is Iceland's (0.195) (Dorfman, 1979). It is worth mentioning that a low Gini coefficient is positive for a country in which the population has buying power. Remarkably, countries such as Austria and Ethiopia have the same Gini coefficient of 0.300. However, this low Gini coefficient means something good for Austria (a homogeneously high living standard), but it means something bad for Ethiopia (a homogeneously low living standard). However it is impressive how flexible this value is, since its application extends to the physics of phase transitions and critical phenomena, and it is used, for example, to characterize condensates that emerge in the flux of particles in counterflowing (Stock et al., 2019).

To adapt the method for calculating the Gini coefficient to this bibliometric context, we proceed as follows: first, rank the members of the population in increasing order of "wealth" (here represented by the h-index), i.e., $h_1 < h_2 < \ldots, h_{n-1} < h_n$. Next, define $\Phi(h_i)$ as the fraction of "bibliometric wealth" associated with the fraction of individuals $f_i = i/n, i = 1, \ldots, n$, which is given by

$$\Phi(h_i) = \frac{\sum_{j=1}^{i} h_j}{\sum_{j=1}^{n} h_j} \tag{8}$$

By applying Equation (8) to each group, the Lorenz curve (Gastwirth, 1972) $(\Phi(h_i), f_i)$ is generated. In a totally fair society (or TPC), we should expect $\Phi(h_i) = i/n$, but in real societies this is not observed. From that, we extend the Lorenz curve concept to describe inequalities in the *h-index* distribution of the scientific population, which is presented for the seven conferences analyzed in this paper in Figure 5.

We can observe that for each group, the area between the Lorenz curve for each conference and the perfect *h-index* distribution represented by the continuous line (identity function $f_i = i/n$) measures the level of inequality in the conference's TPC. This notion can be quantified by Gini statistics or simply by the Gini coefficient. The value of the Gini coefficient is twice the aforementioned area. Theoretically, this coefficient is calculated as in Equation (9):

$$g = 1 - 2\int_0^1 \Phi(h)dh \tag{9}$$

Equation (9) is numerically approximated by a trapezoidal formula, leading to Equation (10):

$$g = 1 - \frac{\Phi(h_0) + \Phi(h_n)}{n} - \frac{2}{n}\sum_{k=1}^{n-1}\Phi(h_k) = 1 - \frac{1}{n}\sum_{k=1}^{n}[\Phi(h_k) + \Phi(h_{k-1})] \tag{10}$$

where $\Phi(h_0) = 0$ and $\Phi(h_n) = 1$ for construction.

A simpler formula (without directly calculating the Lorenz curve) can be used to compute the Gini coefficient. Suppose we have a set of h-indexes for a group: $H = [h_1, h_2, \ldots, h_n]$. An estimator for the Gini coefficient is given by:

$$g = \frac{1}{n\langle h\rangle}\sum_{i=1}^{n}(2i - n - 1)\, h_i, \tag{11}$$

where h_i values are ordered in non-decreasing order, and $\langle h\rangle$ is the mean of h_i.

For a simple example, consider $n = 3$ and $H = [2, 4, 8]$. Then

$$\langle h\rangle = \frac{2+4+8}{3} = \frac{14}{3} \approx 4.6667,$$

and

$$g = \frac{1}{3 \times 4.6667}(-4 + 0 + 16) \approx 0.857.$$

This illustrates that the group has a relatively high inequality in the distribution of h-indices.

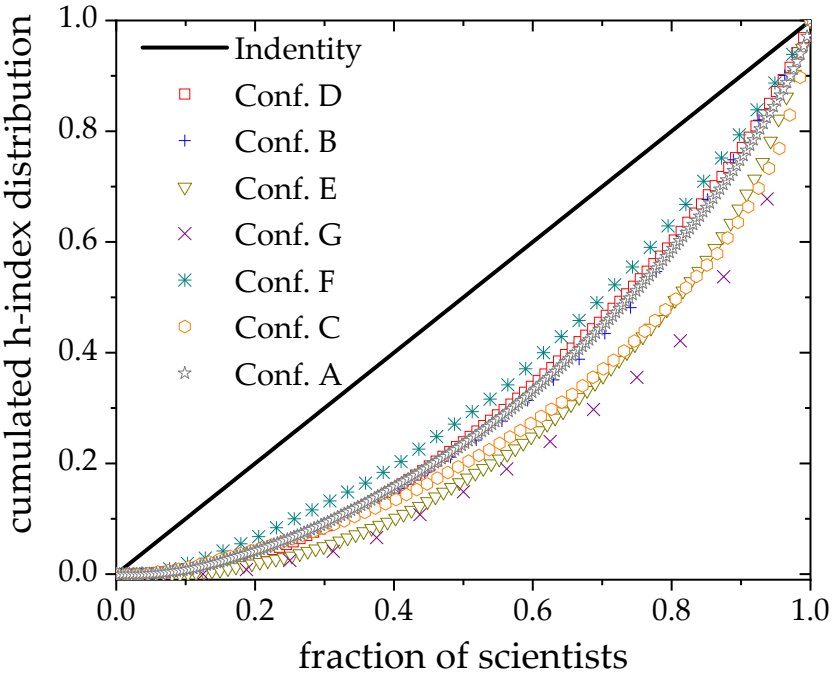


Figure 5. Lorenz curves for the 7 conferences analyzed.

Our proposed method to classify the quality of a group of researchers from the *h-indexes* of its members considers the magnitude of the *h-index* and the level of equality of this *h-index* in the entire TPC population. This new definition, which we call the "*α-index*" is composed of two different quantities: (i) the Gini coefficient of the *h-index* population, and (ii) a definition of the relative *h-index*.

We consider that the *h-index* of a group with n members should be established by the maximum number of members that have an *h-index* equal to or higher than an integer h_{group}, and necessarily the remaining $(n - h_{group})$ members have an *h-index* less than h_{group}.

For example, consider the same three h-index values used in the previous Gini coefficient example: $H = [2, 4, 8]$. Ordered in ascending order, we have $h_1 = 2$, $h_2 = 4$, and $h_3 = 8$. Then, for each i, we compute $n - i + 1$ versus h_i, giving the following points:

$$(n - i + 1, h_i) = (3, 2),\ (2, 4),\ (1, 8).$$

To determine h_{group}, we find the largest integer h, such that at least h members satisfy $h_i \geq h$. In this case,

- $h = 2$ satisfies the condition because there are two members with $h_i \geq 2$.
- $h = 3$ does not satisfy the condition because there are not three members with $h_i \geq 3$.

Therefore, for this group, we have

$$h_{\text{group}} = 2,$$

meaning that there are two members with *h-index* greater than or equal to 2, but fewer than three members with *h-index* ≥ 3.

Up to this point, one may regard h_{group} as the second-order value of the successive *h*-index introduced by Egghe (2008) and Schubert (2007). However, in what follows, we go beyond this notion to formulate our final metrics.

The groups to be compared may have different numbers of members. Thus, to compare different groups, we need to define a relative h_{group} that can be based on the smallest group to be compared. Let us consider the simplest situation: two groups r_1 and r_2 with sizes, respectively, denoted by $|r_1|$ and $|r_2|$, with $|r_1| < |r_2|$. Denoting $H^{(2)} = \{h_1^{(2)}, h_2^{(2)}, \ldots, h_{|r_2|}^{(2)}\}$ as the set of *h-indexes* of members of group r_2, we define the relative h_{group} of r_2 in relation to r_1, over a number of samples (n_{sample}) as the value calculated by Equation (12):

$$h_{group}^{(r_1)}(r_2) = \frac{1}{n_{sample}} \sum_{j=1}^{n_{sample}} h_{group}(S_j^{(r_1)} \subset H^{(2)}) \tag{12}$$

where $S_j^{(r_1)}$ denotes the j-th *h-index* sample of size $|r_1|$ randomly chosen in $H^{(2)}$. This normalization is required because the group r_2 theoretically should have a maximum *h-index* $= |r_2|$, whereas r_1 cannot match that value because it has fewer members. It is important to mention that our definition requires the gathering of samples of "smallest group size" inside of larger groups in a way that groups of different sizes can be compared.

In practice, to find the h_{group} it suffices to plot the function $\psi(h_i) = n - i + 1$ (number of members that have an *h-index* higher than h_i) as a function of h_i and to determine the intercept between $\psi(h_i)$ and the identity function $\phi(h_i) = h_i$ since $h_i < h_{i+1}$, for $i = 1, \ldots, n - 1$ (see Figure 6).

In the case of $m > 2$ groups, a simple procedure is performed:

1. Input: m groups denoted by $r_1, r_2, \ldots, r_m$, and the number of samples (n_{sample}) are required;

2. For every pair of groups (r_{i_x}, r_{i_y}) indexed by $i_x, i_y = 1, \ldots, m$, with $|r_{i_x}| \leq |r_{i_y}|$, samples $S_j^{(i_x)}$ of *h-indexes* of size $|r_{i_x}|$ are randomly selected from $H^{(i_y)} = \{h_1^{(i_y)}, h_2^{(i_y)}, \ldots, h_{|r_{i_y}|}^{(i_y)}\}$, with $j = 1, \ldots, n_{sample}$, and the relative h-group of r_{i_y} with respect to r_{i_x}, i.e., $h_{group}^{(r_{i_x})}(r_{i_y})$, is computed according to Equation (12).

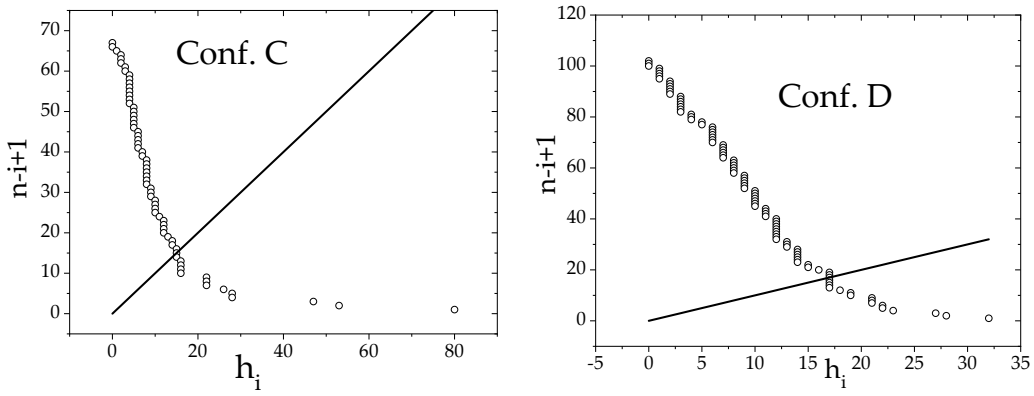


Figure 6. Quantity $n - i + 1$ as a function of h_i, $i = 1, \ldots, n$ for two sample conferences. The intercept with the identity line corresponds to the h_{group}.

From that, a ranking for conferences (or groups) can be established based on their relative h_{group} and the Gini coefficient. Our main proposed function, the *α-index*, is employed to measure the quality of a group l among m groups. Equation (13) defines the *α-index*.

$$\alpha_l = \frac{\frac{(1-g_l)}{(l-k+1)} \sum\limits_{r_k \leq r \leq r_l} h_{group}^{(r)}(r_l)}{\max\limits_{r_k \leq r_j \leq r_l} \{\frac{(1-g_j)}{(j-k+1)} \sum\limits_{r_k \leq r \leq r_j} h_{group}^{(r)}(r_j)\}}, \tag{13}$$

where $k = \arg\min\{|r_i|\}_{i=1}^{m}$ and g_l is the Gini coefficient of group l. The value $0 \leq \alpha_l \leq 1$ measures the quality of a group based on a convenient definition of the *h-index* for groups weighed by the Gini coefficient of members in all groups considered for ranking. The factor g_l works as an amplifier of the relative h_{group}. The smaller the g_l, the more significant the h_{group}.

4. Results–I: Editorial Boards of Conferences

To evaluate the conferences, the first step is to calculate the relative h_{group} of each conference in relation to the smaller conferences. The smallest program committee (Conf. G) has 16 members, while the largest (Conf. A) has 207 members.

For this calculation, we used the simple algorithm described in Section 3, with $n_{\text{sample}} = 1000$, which is more than sufficient. Before proceeding, let us illustrate this point by computing $h_{\text{group}}^{(r_1)}(r_2)$ for the largest group size, $|r_2| = 207$, and for all smaller group sizes $|r_1| = 16, 27, 39, 67, 87, 102$, and 207.

To do so, we vary n_{sample}. We observe that $h_{\text{group}}^{(r_1)}(r_2)$ stabilizes rapidly, with noticeable fluctuations only for small values of $n_{\text{sample}} \lesssim 20$, as shown in Figure 7a. Note that the error bars are very small, of order $O(10^{-2})$, and the standard deviation remains below 1 for $n_{\text{sample}} \gtrsim 100$. We also plot the standard error of the mean (often referred to simply as the standard error), which is used to represent the error bars (rather than the standard deviation), as a function of n_{sample}. As expected, it decreases proportionally to $1/\sqrt{n_{\text{sample}}}$; see Figure 7b.

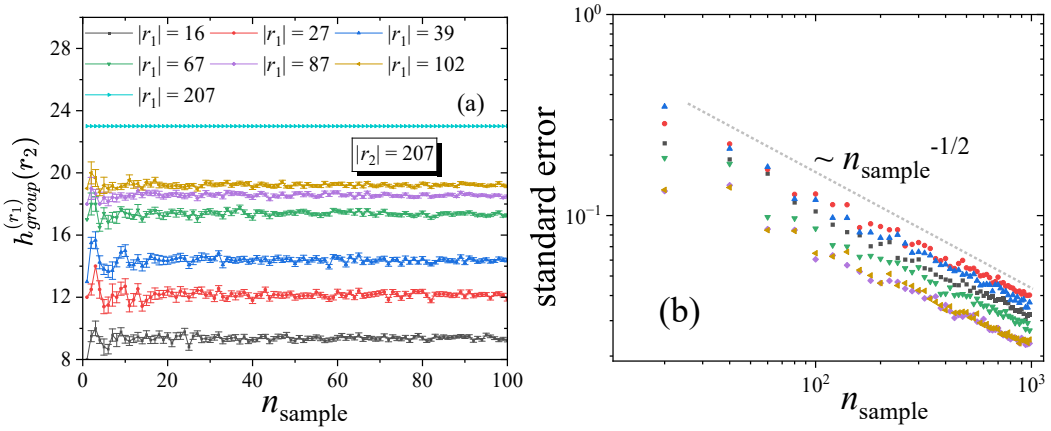


Figure 7. (**a**) $h^{(r_1)}_{\text{group}}(r_2)$ as a function of n_{sample}. (**b**) Standard error of $h^{(r_1)}_{\text{group}}(r_2)$ as a function of n_{sample}.

Table 3 presents our proposed ranking of the conferences according to the *α-index*. The table also shows the average *h*-index of the members of each conference. The h_{group} values reported in this table correspond to $h^{(r_l)}_{group}(r_l)$, i.e., computed using the entire population of members in each conference, representing the maximum h_{group}.

Table 3. Average *h-index*, Gini coefficient, maximum h_{group}, and *α*-index for the 7 conferences analyzed.

Conference	$\langle h_{index} \rangle$	Gini Coef.	h_{group}	*α*-Index
Conf. A (A)	12.78 ± 0.65	0.377	23	1.000
Conf. D (A)	10.10 ± 0.66	0.367	17	0.820
Conf. C (B)	11.63 ± 1.55	0.462	15	0.700
Conf. B (C)	11.92 ± 1.60	0.381	12	0.652
Conf. E (B)	08.07 ± 0.83	0.487	14	0.648
Conf. F (A)	07.94 ± 0.69	0.303	10	0.570
Conf. G (C)	07.56 ± 2.39	0.548	6	0.380

Our results show a divergent ranking from the one that would be established by the simple calculation of the average of the *h-indexes*. Our *α-index* shows the need for the inclusion of the Gini coefficient or another homogeneity parameter in the analysis of the quality of conferences. Many conferences have a high h_{group} due to just a small fraction of the TPC. The Gini coefficient shows how representative the computed h_{group} is. A low g denotes that the conference has a robust h_{group}. Furthermore, it means that any smaller sample collected from the group should have the same h_{group}, making it independent from the sample. Conferences with a high Gini coefficient have discrepant TPC members, which is a sign of questionable quality.

The ranking of conferences according to the *α-index* differs from the ranking by CAPES, the research funding agency, in two cases. Conf. F is given a lower ranking according to the *α-index*. In fact, in a later assessment, CAPES re-evaluated subsequent editions of this conference, placing it at a lower rank. The second discrepancy was for Conf. B, which ranked higher according to the *α-index*. CAPES uses a minimum number of editions as an attribute to determine the quality of the conferences, and this may have led to the incorrect ranking of Conf. B. Our approach does not depend on the number of editions as it analyzes the h-index distribution for a specific edition of a conference, and this is one of its main advantages.

Another characteristic of the *α-index* is that the relative ordering of the groups would remain the same if only a subset of the conferences had been compared. Furthermore, if we do pairwise comparisons, for example, between conference X and Y, and find that X

is better than Y, then comparing Y to Z, we find that Y is better than Z. By transitivity, a comparison between X and Z would result in X having a higher α-*index* than Z.

The key question here concerns whether this ranking could replace the classification established by the Brazilian agency CAPES for conferences. To address this, we calculated the Spearman correlation coefficient ρ, which measures the monotonic relationship between two variables; the Kendall coefficient (τ), a non-parametric alternative suitable for small samples and ties; and finally, the Kruskal–Wallis test, another non-parametric approach, to evaluate the relationship between the α-index and the assigned quality levels: 3 for A-type conferences, 2 for B-type, and 1 for C-type. For comparison, we also computed the same correlations using the simple average $\langle h_{index} \rangle$. The results are summarized in Table 4.

Table 4. Spearman and Kendall correlation coefficients, and Kruskal–Wallis test results for the average h-index and the α-index. These results support that the α-index performs better, showing higher correlations and smaller p-values in the Kruskal–Wallis test.

Metrics	Spearman	Kendall	Kruskall-Walis
$\langle h_{index} \rangle$	0.170	0.109	0.179, $p = 0.915$
α-index	0.624	0.546	3.179, $p = 0.204$

These results corroborate that the α-index provides a better assessment, as it exhibits higher correlations and smaller p-values in the Kruskal–Wallis test. It is important to emphasize that the index is not intended to exactly reproduce the classification trends established by CAPES, but rather to highlight possible distortions in the current system, which may not fully account for group homogeneity and the relative average h-group. In other words, the goal is to propose a pragmatic alternative to replace subjective criteria that may compromise the evaluation process. This consideration is essential to ensure that larger groups are not privileged to the detriment of smaller but higher-quality ones.

Finally, other interesting instances for our approach could be easily experimented. For example, one could consider not only the *h-index* of the TPC members but also the *h-indexes* of the authors who have published papers in the conference. The difficulty here is the greater quantity of data required and its pre-processing. In the following section we show results for postgradute programs as an alternative validation of the methodology.

5. Results—II: Additional Validation Through Postgraduate Programs

To test our methodology in a different context, we also collected data from members of several postgraduate programs in Physics in Brazil. As in the case of conferences, the data refer to an earlier period (accessed in 2010) in the Google Scholar service (http://scholar.google.com), well before the COVID-19 pandemic, thereby avoiding any suggestion of bias. The sample includes nine postgraduate programs. According to CAPES's classification, levels 7 and 6 correspond to excellent programs, level 5 corresponds to good programs, and levels 4 and 3 are considered regular. For confidentiality reasons, we do not disclose the names of the programs or the exact year of data collection. As in the previous analysis, the data source was Google Scholar.

Table 5 summarizes the data corresponding to the postgraduate programs. Unlike the case of conference editorial boards, here we computed all values of $h^{(r)}_{group}(r_l)$ for $r_k \leq r \leq r_l$ in each program. The values in bold indicate the maximum $h^{(r_l)}_{group}(r_l)$ for each group $l = 1, \ldots, 9$, corresponding, respectively, to programs P1, P2, ..., P9. Each column contains the set of values $h^{(r)}_{group}(r_l)$ for $r_k \leq r \leq r_l$, while the last row shows the average $A(r_l) = \frac{1}{(l-k+1)} \sum_{r=r_k}^{r_l} h^{(r)}_{group}(r_l)$, which is later used to compute the α-index according to Equation (13).

Table 5. Comparative results for selected Brazilian postgraduate programs in Physics. The columns present the relative h-indices, from the maximum value (in bold) down to the minimum (corresponding to the size of the smallest group). The last row reports the average of the relative h-groups $A(r_l)$, which is used to compute the α-index. Complete this table by adding, in the last three columns, the group size, the Gini coefficient, and finally the α-index.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**Size**	g	α
P1 (7)	**18.00**									85	0.258	1.000
P2 (7)	14.95	**14.00**								56	0.326	0.802
P3 (7)	14.74	14.00	**14.00**							53	0.256	0.872
P4 (5)	12.41	12.29	11.69	**11.00**						28	0.278	0.744
P5 (6)	11.99	11.42	11.24	10.39	**13.00**					24	0.283	0.888
P6 (5)	11.85	11.16	11.10	10.23	12.45	**11.00**				23	0.217	0.846
P7 (4)	11.60	10.57	10.85	9.89	12.09	10.75	**8.00**			21	0.265	0.571
P8 (4)	11.03	9.64	10.42	9.37	11.68	10.14	7.60	**9.00**		18	0.277	0.619
P9 (4)	10.37	8.75	9.81	8.76	10.48	9.81	6.89	8.26	**9.00**	15	0.180	0.765
$A(r_l)$	12.99	11.48	11.30	9.93	11.94	10.43	7.50	8.26	9.00			

The table contains several pieces of information—the number in parentheses after each program denotes its CAPES level (4, 5, 6, or 7). The three rightmost columns show, respectively, the number of members in each program, the corresponding Gini coefficient (also used in the computation of the α-index), and finally the value of the α-index itself.

We calculated the Spearman and Kendall correlation coefficients between the CAPES level of each program and our α-index. These tests yielded $\rho = 0.788$ and $\tau = 0.650$, with very low p-values of $p = 0.012$ and $p = 0.127$, respectively. The Kruskal–Wallis test provided consistent results, with a statistic of $K = 5.711$ and $p = 0.127$.

The good correlations suggest that CAPES's classification captures the general structure of program quality; however, some aspects deserve further attention, particularly when comparing smaller programs to larger ones. Our metric indicates that such evaluations could be improved by incorporating additional quantitative details—such as group homogeneity and representative sampling—in order not to exclude smaller programs, while simultaneously avoiding penalizing larger ones, provided these are not composed mainly of a few "hot" researchers surrounded by many less productive members.

6. Summary and Conclusions

This paper proposes a new method for classifying research groups across any scientific field. To motivate the approach, we begin with a preliminary pedagogical analysis that describes in detail the statistical properties of groups of researchers, establishing some universal bibliometric stylized facts based on the program committees of scientific conferences.

Next, we formally introduce a quantity that combines the concepts of homogeneity (via the Gini coefficient) and magnitude (through the relative h_{group}), which we call the α–index. By analyzing both normal and non-normal groups of researchers, we establish a ranking for the seven conferences examined.

When applied to the CAPES classification scheme, our method reveals a strong statistical correlation, indicating that a fair assessment should rely on more than a high average *h-index*. Features such as group homogeneity, together with a reasonable average *h-index*, should also be incorporated as criteria to achieve a more balanced and conceptually sound evaluation.

Our conclusions were supported by correlation tests using Spearman, Kendall, and Kruskal–Wallis statistics. Although the correlations are high, it is important to emphasize that our work is primarily recommendatory, illustrating that additional refinements are needed to improve comparisons between smaller and larger groups.

To further demonstrate the versatility of the α–index, we also applied it to classify nine postgraduate programs in Physics in Brazil, finding strong agreement with the classification established by CAPES. In this second analysis, the same statistical tests previously applied to the conferences yielded extremely low p-values, providing additional evidence for the robustness of our ranking methodology.

It is important to note that the definition of the group h–index (simply h_{group}), used in the construction of the α–index, relies first on the concept of the relative h_{group}. This quantity corresponds to the h_{group} that a larger group would have if it were reduced to the size of the smaller group. An average over all smaller groups, as well as over its own size, is then considered. This approach captures the superiority of a larger group only when it is truly a well-composed program, without disproportionately favoring group size.

It is important to mention that numerous studies have successfully examined the evaluation of research groups (see, for example, Abramo and D'Angelo (2014); Berche et al. (2016); Biagetti (2022); Fernandes and Lopes (2018); Rosas et al. (2011); Thaule and Strand (2018); van Raan (2012); Waltman and van Eck (2015)), reinforcing the relevance of this topic. A comparative analysis among these methods would also be valuable, especially because many of them rely on qualitative assessments that require human supervision.

It also deserves further investigation, since the method should be applied to characterize the quality of groups of scientists using other databases, such as ISI–JCR (Garfield, 2009). In particular, we should ask whether one should expect systematic differences or distortions when different databases are considered.

Naturally, some criticisms and warnings must be raised here. Although Google Scholar offers broad coverage—including regional journals, multiple languages, and diverse document types—its use in bibliometric analyses demands caution. Its coverage may fluctuate over time, with documents occasionally disappearing from the index, thus affecting the reproducibility of citation counts (Martín-Martín & Delgado López-Cózar, 2021). Moreover, the lack of transparent indexing criteria, heterogeneous metadata quality, and the presence of duplicate or inconsistently parsed records introduce noise into the dataset (Halevi et al., 2017). The limited search interface—with restrictions on query complexity and filtering—also reduces the ability to replicate searches precisely (Haddaway et al., 2015). For these reasons, metrics derived from Google Scholar should be interpreted with caution and, whenever possible, cross-checked with more controlled databases such as Web of Science or Scopus (Martín-Martín et al., 2018).

Despite these criticisms, we note that some studies suggest that the h-index follows a universal distribution (da Silva et al., 2012), independent of the database used. The main hypothesis of the present study is that, regardless of the database, such distortions appear to be "democratically" distributed, and cross-database comparisons tend to reflect primarily a scale effect, often yielding similar rankings within the same source.

Author Contributions: Conceptualization, R.d.S., J.P.M.d.O. and V.M.; Methodology, R.d.S., J.P.M.d.O. and V.M.; Software, R.d.S., J.P.M.d.O. and V.M.; Validation, R.d.S., J.P.M.d.O. and V.M.; Formal analysis, R.d.S., J.P.M.d.O. and V.M.; Investigation, R.d.S., J.P.M.d.O. and V.M.; Resources, R.d.S., J.P.M.d.O. and V.M.; Data curation, R.d.S., J.P.M.d.O. and V.M.; Writing—original draft, R.d.S., J.P.M.d.O. and V.M.; Writing—review & editing, R.d.S., J.P.M.d.O. and V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Council for Scientific and Technological Development grant numbers 304575/2022-4, and JPMO under grants 306695/2022-7, 405973/2021-7, and 402086/2023-6

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*(2), 1129–1144. [CrossRef]

Almeida Guimarães, J., & Chaves Edler de Almeida, E. (2012). Quality assurance of post-graduate education: The case of CAPES, the Brazilian agency for support and evaluation of graduate education. *Higher Learning Research Communications*, *2*(3), 3–11. [CrossRef]

Batista, P. D., Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, *68*(1), 179–189. [CrossRef]

Berche, B., Holovatch, Y., Kenna, R., & Mryglod, O. (2016). Academic research groups: Evaluation of their quality and quality of their evaluation. In *Journal of physics: Conference series* (Vol. 681, p. 012004). IOP Publishing. [CrossRef]

Biagetti, M. T. (2022). From bibliometric indicators to qualitative analysis: The valorisation of peer-review in the evaluation of scientific research. *Bibliothecae.It*, *11*(2), 331–349. [CrossRef]

Biagioli, M., & Lippman, A. (Eds.). (2020). *Gaming the metrics: Misconduct and manipulation in academic research*. MIT Press. [CrossRef]

Braun, T. (2004). Keeping the gates of science journals: Gatekeeping indicators of national performance in the sciences. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 95–114). Kluwer Academic Publishers.

da Silva, R., Kalil, F., de Oliveira, J. P. M., & Martinez, A. S. (2012). Universality in bibliometrics. *Physica A: Statistical Mechanics and Its Applications*, *391*(6), 2119–2128. [CrossRef]

da Silva, R., Lamb, L. C., & Barbosa, M. C. (2016). Universality, correlations, and rankings in the Brazilian universities national admission examinations. *Physica A: Statistical Mechanics and its Applications*, *457*, 295–306. [CrossRef]

Dorfman, R. (1979). A Formula for the gini coefficient. *The Review of Economics and Statistics*, *61*(1), 146–149. [CrossRef]

Egghe, L. (2008). Modelling successive h-indices. *Scientometrics*, *77*(3), 377–387. [CrossRef]

Fernandes, G., & Lopes, M. (2018). Research evaluation, bibliometric indicators and impact on knowledge development. *The IUP Journal of Knowledge Management*, *XV*(1), 12.

Garfield, E. (2009). The evolution of the science citation index. *Contributions to Science*, *5*(1), 63–70.

Gastwirth, J. L. (1972). The estimation of the lorenz curve and gini index. *The Review of Economics and Statistics*, *54*(3), 306–316. [CrossRef]

Giddings, J. C., & Eyring, H. (1954). A molecular dynamic theory of chromatography. *The Journal of Chemical Physics*, *22*(3), 416–421. [CrossRef]

Gini, C. (1921). Measurement of inequality and incomes. *The Economic Journal*, *31*, 124–126. [CrossRef]

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of google scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, *10*(9), e0138237. [CrossRef] [PubMed]

Halevi, G., Moed, H. F., & Bar-Ilan, J. (2017). Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics*, *11*(3), 823–834. [CrossRef]

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. [CrossRef]

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. [CrossRef]

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: Fat tails with characteristic scales. *The European Physical Journal B*, *2*(4), 525–539. [CrossRef]

Lima, S. R. A. d. (2025). Capes guidelines and procedures for evaluating *stricto sensu* postgraduate programmes in Brazil. *Revista da Tulha*, *11*(1), 1–22. [CrossRef]

Lopes, G. R., da Silva, R., & de Oliveira, J. P. M. (2011, May 25–27). *Applying Gini coefficient to quantify scientific collaboration in researchers network*. Proceedings of the International Conference on Web Intelligence, Mining and Semantics, Sogndal Norway. [CrossRef]

Martín-Martín, A., & Delgado López-Cózar, E. (2021). Large coverage fluctuations in Google Scholar: A case study. *arXiv*, arXiv:2102.07571. [CrossRef]

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. [CrossRef]

Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, *66*(10), 1988–2002. [CrossRef]

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. [CrossRef]

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*, 5200–5205. [CrossRef]

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B*, *4*(2), 131–134. [CrossRef]

Rosas, S. R., Kagan, J. M., Schouten, J. T., Slack, P. A., & Trochim, W. M. K. (2011). Evaluating research and impact: A bibliometric analysis of research by the NIH/NIAID HIV/AIDS clinical trials networks. *PLoS ONE*, *6*(3), e17428. [CrossRef]

Schubert, A. (2007). Successive h-indices. *Scientometrics*, *70*(1), 201–205. [CrossRef]

Silva, F. N., Rodrigues, F. A., Oliveira, O. N., & da F. Costa, L. (2013). Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics*, *7*(2), 469–477. [CrossRef]

Stock, E. V., da Silva, R., & da Cunha, C. R. (2019). Numerical study of condensation in a Fermi-like model of counterflowing particles via Gini coefficient. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(8), 083208. [CrossRef]

Thaule, J., & Strand, M. (2018, June 17–21). *Evaluating evaluative bibliometrics: A case study of two research groups*. Proceedings of the IATUL Conferences, Oslo, Norway.

Tsallis, C., & de Albuquerque, M. P. (2000). Are citations of scientific papers a case of nonextensivity? *European Physical Journal B*, *13*(4), 777–780. [CrossRef]

van Raan, A. F. J. (2012). Properties of journal impact in relation to bibliometric research group performance indicators. *Scientometrics*, *92*(2), 457–469. [CrossRef] [PubMed]

Waltman, L., & van Eck, N. J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. [CrossRef]

Waltman, L., & van Eck, N. J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, *7*(4), 833–849. [CrossRef]

Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*(4), 872–894. [CrossRef]

Zhuang, Z., Elmacioglu, E., Lee, D., & Giles, C. L. (2007, June 18–23). *Measuring conference quality by mining program committee characteristics*. Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries (pp. 135–144), Vancouver, BC, Canada.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.