



OPEN ACCESS

EDITED BY

Alex Hansen,
Norwegian University of Science and
Technology, Norway

REVIEWED BY

Sitangshu Bikas Santra,
Indian Institute of Technology
Guwahati, India
Francisco Wellington Lima,
Federal University of Piauí, Brazil
Tomonari Sei,
The University of Tokyo, Japan

*CORRESPONDENCE

Manipushpak Mitra,
mmitra@isical.ac.in

SPECIALTY SECTION

This article was submitted to
Interdisciplinary Physics,
a section of the journal
Frontiers in Physics

RECEIVED 15 August 2022

ACCEPTED 24 October 2022

PUBLISHED 17 November 2022

CITATION

Ghosh A, Chakrabarti BK, Ram DRS,
Mitra M, Maiti R, Biswas S and Banerjee S
(2022), Scaling behavior of the Hirsch
index for failure avalanches, percolation
clusters, and paper citations.
Front. Phys. 10:1019744.
doi: 10.3389/fphy.2022.1019744

COPYRIGHT

© 2022 Ghosh, Chakrabarti, Ram, Mitra,
Maiti, Biswas and Banerjee. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Scaling behavior of the Hirsch index for failure avalanches, percolation clusters, and paper citations

Asim Ghosh¹, Bikas K. Chakrabarti^{2,3}, Dachehalli R. S. Ram⁴,
Manipushpak Mitra^{3*}, Raju Maiti³, Soumyajyoti Biswas⁴ and
Suchismita Banerjee³

¹Department of Physics, Raghunathpur College, Raghunathpur, India, ²Department of Physics, Saha Institute of Nuclear Physics, Kolkata, India, ³Economic Research Unit, Indian Statistical Institute, Kolkata, India, ⁴Department of Physics, SRM University, Amaravathi, Andhra Pradesh, India

A popular measure for citation inequalities of individual scientists has been the Hirsch index (h). If for any scientist the number n_c of citations is plotted against the serial number n_p of the papers having those many citations (when the papers are ordered from the highest cited to the lowest), then h corresponds to the nearest lower integer value of n_p below the fixed point of the non-linear citation function (or given by $n_c = h = n_p$ if both n_p and n_c are a dense set of integers near the h value). The same index can be estimated (from $h = s = n_s$) for the avalanche or cluster of size (s) distributions (n_s) in the elastic fiber bundle or percolation models. Another such inequality index called the Kolkata index (k) says that $(1 - k)$ fraction of papers attract k fraction of citations ($k = 0.80$ corresponds to the 80–20 law of Pareto). We find, for stress (σ), the lattice occupation probability (p) or the Kolkata Index (k) near the bundle failure threshold (σ_c) or percolation threshold (p_c) or the critical value of the Kolkata Index k_c a good fit to Widom–Stauffer like scaling $h/[\sqrt{N}/\log N] = f(\sqrt{N}[\sigma_c - \sigma]^\alpha)$, $h/[\sqrt{N}/\log N] = f(\sqrt{N}|p_c - p|^\alpha)$ or $h/[\sqrt{N_c}/\log N_c] = f(\sqrt{N_c}|k_c - k|^\alpha)$, respectively, with the asymptotically defined scaling function f , for systems of size N (total number of fibers or lattice sites) or N_c (total number of citations), and α denoting the appropriate scaling exponent. We also show that if the number (N_m) of members of parliaments or national assemblies of different countries (with population N) is identified as their respective h – indexes, then the data fit the scaling relation $N_m \sim \sqrt{N}/\log N$, resolving a major recent controversy.

KEYWORDS

Hirsch index, Kolkata index, percolation model, fiber bundle models, paper citation

1 Introduction

Monotonically and nonlinearly decaying inequality functions are ubiquitous. When the number (n_p) of papers by any author (or, for that matter, an institution) is arranged according to the number (n_c) of the citations they received, the citation inequality function becomes a monotonically decaying one (see e.g. [1]). The same is true for avalanches in material failure or in earthquakes (see e.g. [2]), cluster size distributions in percolation problems (see e.g. [3]), etc., where the number (n_s) of avalanches or clusters (giving the size inequality function) decreases monotonically and nonlinearly with the size (s) of the avalanche or cluster. Large avalanches, strong quakes, or big size clusters come or occur in small numbers, while the smaller or weaker the avalanches, quakes, or clusters, the larger is their abundance in occurrence. How does one statistically measure these inequalities in occurrence frequencies of citation numbers or avalanche or cluster sizes? Obviously, the corresponding distribution functions for inequalities in citations or sizes would contain the entire statistics. However, they are not convenient to handle. One can consider the citation number of the best cited paper (as sometimes carried out for some unique awards, etc.) or study the statistical (self-similar) structure of the biggest avalanche or the largest (percolating) cluster (as in statistical physics [3]). Hirsch proposed [1] (Figure 1) an index to

measure these inequalities, by locating the fixed point of the nonlinear inequality function. The Hirsch index (h) corresponds to the citation number or occurrence frequency which is commensurate in magnitude with the number of publications or avalanche cluster sizes.

Systems near their critical points, self-organized or tuned, have mostly been studied in the self-similar limit of their divergent correlations (see e.g. [2, 3]). The corresponding critical exponent values arising out of such self-similarities have helped classify vastly different physical systems based on their symmetries, dimensionality, and such broad qualifiers. One powerful tool has been the scaling relations among the critical exponents that helped build an interconnected and precise relation between experimental observables in such systems near criticality. In this work, however, we focus on quantifying the response of near-critical systems through the corresponding inequality statistics (for example, in n_c versus n_p of citations or in n_s versus s for avalanches or clusters). Specifically, we measure the widely used Hirsch index (h), which in effect, gives a measure of a size that is commensurate with its relative abundance. It turns out to be possibly even more robust than the critical behavior (characterized by a set of exponents). We demonstrate this by choosing a wide variety of systems, namely, citation statistics, percolation cluster statistics, and avalanche statistics, in fiber bundle models (FBMs) and even in the statistics of parliament sizes in different countries of the world. They differ in their dimensionality (two dimensions for the percolation model studied here, mean-field for the fracture model, and possibly small-world networks for citation and parliament statistics). They further differ in their measurement variables and their size distribution statistics, making them widely different in terms of their prominently apparent features.

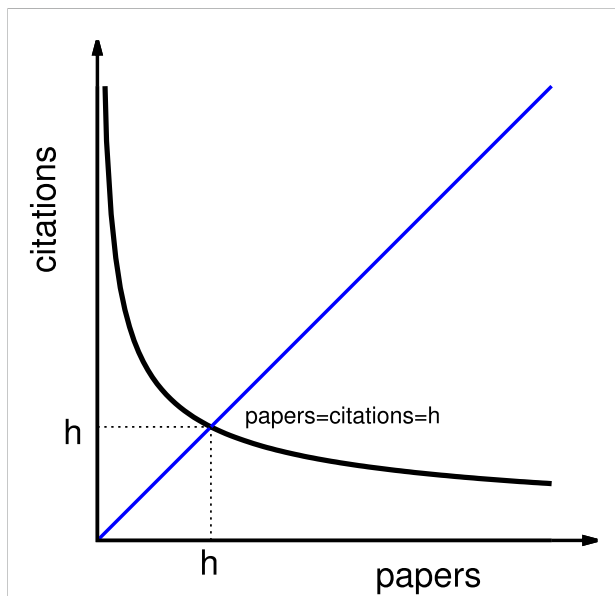


FIGURE 1
Schematic drawing of the citation function of a typical scientist. The h -index (an integer) is given by the lower value of the paper serial number below the fixed point value (the intersection point of the 45° line from the origin), when the papers are ordered from the highest cited to the lowest cited one. When the citation number and the serial number of the papers (having those citations) are both sequential integers near the fixed point, the citation number equals the number of papers and both become h . Similar will be the case where citations are replaced by the failure of the material avalanche sizes (or cluster sizes), and the paper numbers are replaced by the number of such avalanches (or of the clusters in pre- or post-percolating systems).

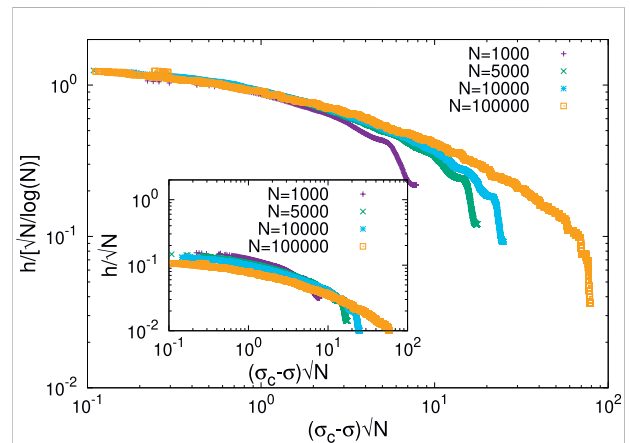


FIGURE 2
Finite size scaling analysis of the fiber bundle model h -index (from avalanche size distribution). The scaling fit to the Widom–Staufer relations (Eq. 1) is obtained by making all the data points of different sizes ($N = 1000, 5000, 10000$, and 100000) and stress level (different σ values) collapse together. It turns out that the scaling form (Eq. 1b) with the exponent $\alpha = 1$ gives good data collapse. The inset shows that the data collapse clearly gets worsened by dropping the $\log N$ term in the scaling relation (Eq. 1b).

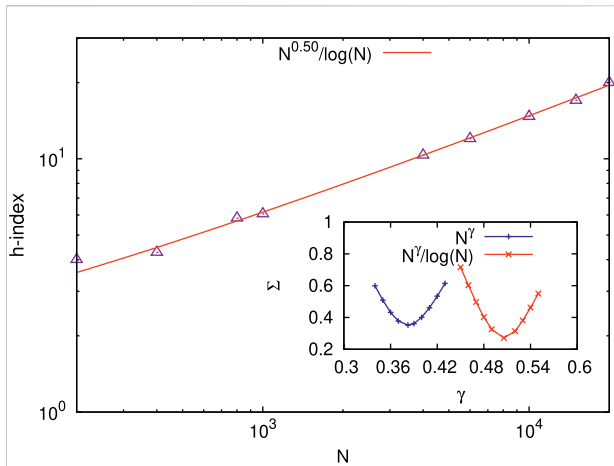


FIGURE 3 Scaling behavior for the average h -index (in the range $4 \leq h \leq 20$) at the breaking point ($\sigma = \sigma_c$) of the bundles with the total number N (in the range $200 \leq N \leq 20000$) of fibers in the equal-load-sharing FBM (with uniform distribution of fiber breaking thresholds) considered here (cf [9]). The figure shows the best fit of h to $\sqrt{N}/\log N$. The inset shows statistical deviations $\Sigma = \sqrt{\sum_i (h_d(i) - h_y(i))^2 / \Sigma_i} / 1$ to the scaling forms $h \sim N^\gamma$ and $h \sim N^\gamma / \log(N)$ with respect to γ , where i represents the individual (random) realization of the FBM with the corresponding $h (= h_d)$ value obtained from the simulation result and the scaling fit (h_y) from the aforementioned scaling relations. The inset shows that the statistical deviation becomes minimum for the scaling with the log correction term with the appropriate value of γ near $1/2$.

However, we show here that in spite of their obvious differences, the scaling behavior of the Hirsch index shows remarkable universality.

How does the h -index scale with a total number of publications (N_p) by the author (institution) or the total number (N_c) of citations received by the author (institution)? Young [4] suggested analytically that the h -index value should scale with the total number of citations $N_c (= \sum n_c)$ as $\sqrt{N_c}$ asymptotically. In [5], a brief analysis of the Google Scholar data indicated independently that $h \sim \sqrt{N_p}$ for large values of N_p which corresponds to the highest value of n_p , the total number of publications by the author. It should be noted that if both these relationships are valid then, statistically speaking, for a prolific author, the total number of citations would be linearly proportional to the total number of papers published, and the proportionality constant would perhaps be determined by the size of the existing author network in the subject (suggesting an effective Dunbar number [7, 8], for the community of authors).

In a recent Monte Carlo study [9], on the avalanche sizes and their numbers in the fiber bundle models (FBMs) of material failure (see e.g. [2, 10]) due to increasing stress on such bundles, the numerical analysis of the data for the nonlinearly decaying numbers of avalanches with their sizes (or released elastic energies) suggested $h \sim \sqrt{N}/\log N$.

Our numerical study on avalanche size distributions in FBMs (for stress or load per fiber σ less than its global failure stress σ_c) of cluster size distributions for the lattice occupation concentration p near the percolation point p_c (for p both below and above p_c) in the percolating system and the previous analysis [5] of citation

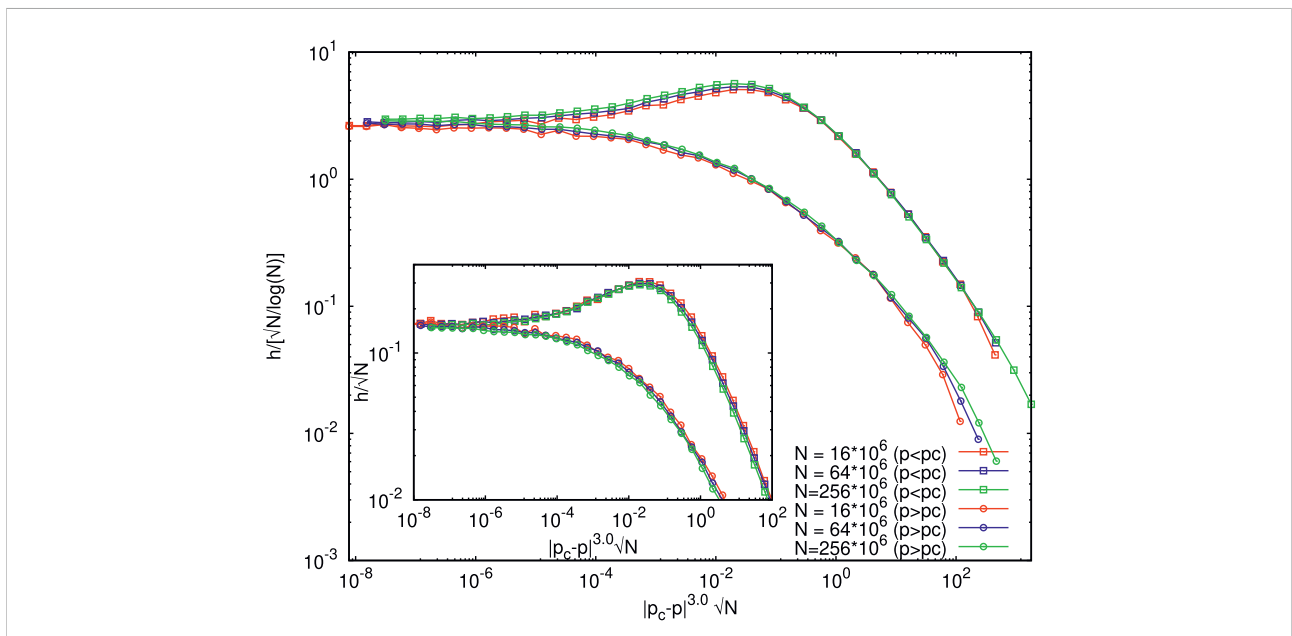


FIGURE 4 Finite size scaling analysis of the 2D site percolation h -index (from the cluster size distribution). The scaling fit to the Widom–Stauffer relations (Eq. 1) is obtained by making all the data points of different sizes ($N = 4000 \times 4000, 8000 \times 8000$, and 16000×16000) and concentrations (different p values, $p_c = 0.5927$) collapse together. It turns out that the scaling form (Eq. 1b) with the exponent $\alpha = 3$ gives good data collapse. The inset shows that the data collapse seems to get worsened by dropping the $\log N$ term in the scaling relation (Eq. 1b).

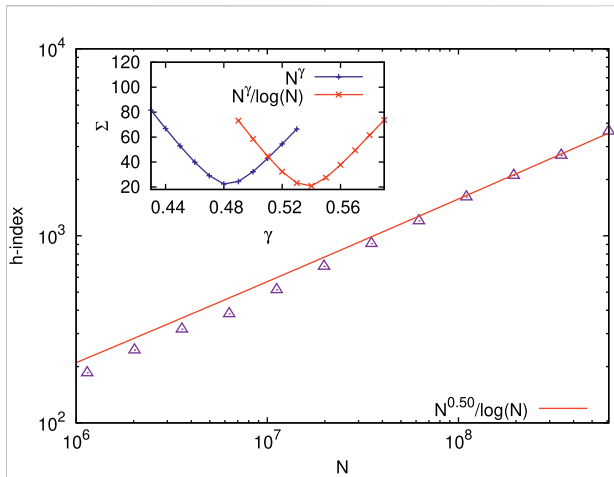


FIGURE 5
Scaling behavior for the h -index (in the range $185 \leq h \leq 4421$) for cluster size distributions at the percolation point (with size N on the square lattice; $10^6 < N < 10^9$). The figure shows the best fit of h to $\sqrt{N}/\log N$. The inset shows the statistical deviations $\Sigma = \sqrt{\sum_i (h_d(i) - h_\gamma(i))^2 / \Sigma_i}$ to the scaling forms $h \sim N^\gamma$ and $h \sim N^\gamma/\log(N)$ with respect to γ , where i represents the individual (random) realization of the percolating system with the corresponding h ($= h_d$) value obtained from the simulation result and the scaling fit (h_γ) from the aforementioned scaling relations. The inset shows that the statistical deviation becomes minimum for the scaling with the log correction term, although with a value of γ somewhat above $1/2$.

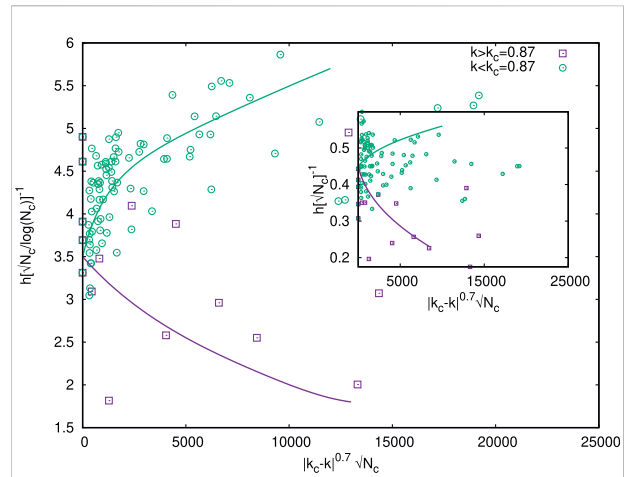


FIGURE 6
Finite size (N_c) scaling analysis of the citation distributions of 100 individual scientists (Table 1 of Ref. [5]). The scaling fit to the Widom–Stauffer relation (Eq. 1c) is obtained by fitting all the data for their h -index and the N_c values of total citations of the publications by individual scientists and their corresponding Kolkata index values k near the critical value $k_c \approx 0.87$ [5]. It turns out that the scaling form (Eq. 1c) with the exponent $\alpha = 0.7$ gives a good data collapse. The inset shows that the data collapse seems to get worsened by dropping the $\log N$ term in the scaling relation (Eq. 1c), although at the critical point $k = k_c$, one gets $h = a\sqrt{N_c}$, with $a = 0.45 \pm 0.10$, which includes the analytical prediction [4] that $a = 0.54$ in the large N_c limit. The continuous lines (in both the main figure and the inset) represent the scaling functions for $k < k_c$ (green) and $k > k_c$ (violet) and are drawn for a guide to the eye.

distributions of scientists with the Kolkata index k (see e. g. [6] for a review, giving the fraction k of citations/wealth attracted/possessed by $1 - k$ fraction of publications/people) near (both above and below) the threshold point k_c ($= 0.86$) both show an excellent fit to a Widom–Stauffer like scaling relation between the Hirsch index (h) and the system size N (or individual’s total citation size N_c), following Widom scaling for the free energy away from the critical point and the subsequent Stauffer scaling [3] for the number of a particular sized cluster, identified here as the equivalent Hirsch index, at and away from the percolation point):

$$\frac{h}{[\sqrt{N}/\log N]} = f(\sqrt{N} [\sigma_c - \sigma]^\alpha), \tag{1a}$$

$$\frac{h}{[\sqrt{N}/\log N]} = f(\sqrt{N} |p_c - p|^\alpha), \tag{1b}$$

$$\frac{h}{[\sqrt{N}/\log N]} = f(\sqrt{N}_c |k_c - k|^\alpha), \tag{1c}$$

with the asymptotically well-defined finite size scaling function f ($f(x) = \text{constant}$ at $x = 0$ and f remaining continuous and finite as x approaches infinity) for systems of size N (total number of fibers or lattice sites/bonds) or size N_c (total number of citations of all the publications by an individual scientist) and α denoting the appropriate scaling exponent.

Traditionally, the Hirsch index h for different authors has been fitted [4] to the scaling form \sqrt{N}_c . For cases where the Kolkata index values of the authors are not known, we simply fitted, following the scaling relation (Eq. 1), to the form appropriate for the critical point:

$$h \sim \sqrt{N}_c / \log N_c \tag{2a}$$

$$\sim \sqrt{N}_p / \log N_p. \tag{2b}$$

We also show that if the number (N_m) of members of the parliaments or national assemblies of different countries (with the corresponding population denoted by N) is identified as those countries’ respective h indices, then the data fit well to the scaling relation $N_m \sim \sqrt{N}/\log N$. This helps comprehend the discrepancies (c.f [11]) with the \sqrt{N} relationship, reported in a very recent analysis [12], resolving a recent major controversy.

2 Data analysis and numerical studies

2.1 Failure avalanches in the fiber bundle model (FBM)

The fiber bundle model (FBM) is a generic model for failure of disordered solids. An ensemble of N fibers is a set

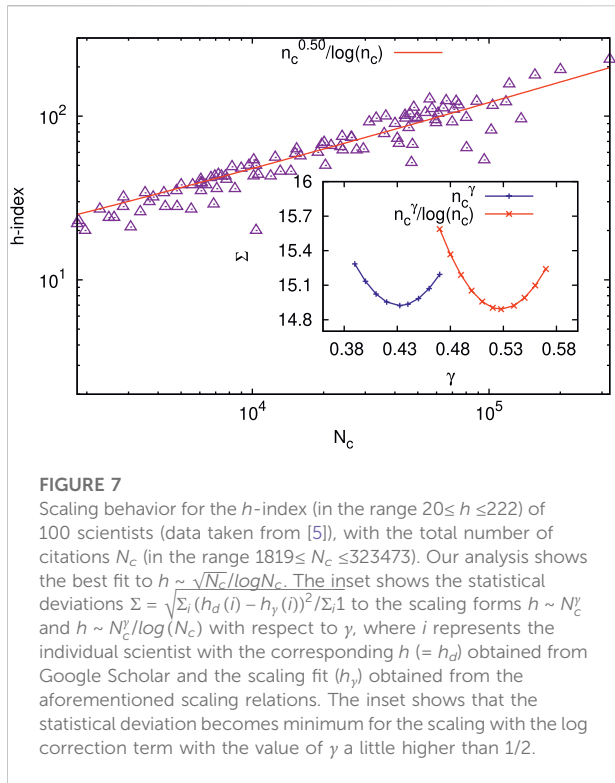


FIGURE 7
Scaling behavior for the h -index (in the range $20 \leq h \leq 222$) of 100 scientists (data taken from [5]), with the total number of citations N_c (in the range $1819 \leq N_c \leq 323473$). Our analysis shows the best fit to $h \sim \sqrt{N_c}/\log N_c$. The inset shows the statistical deviations $\Sigma = \sqrt{\sum_i (h_d(i) - h_\gamma(i))^2 / \sum_i 1}$ to the scaling forms $h \sim N_c^\gamma$ and $h \sim N_c^\gamma / \log(N_c)$ with respect to γ , where i represents the individual scientist with the corresponding $h (= h_d)$ obtained from Google Scholar and the scaling fit (h_γ) obtained from the aforementioned scaling relations. The inset shows that the statistical deviation becomes minimum for the scaling with the log correction term with the value of γ a little higher than $1/2$.

between two rigid parallel plates, and a load is applied on the bottom plate. Each fiber is linear elastic with the same elastic constant and has a failure threshold selected randomly from a distribution. This failure threshold is the source of the disorder and non-linearity in the otherwise linear model.

When a small load (W) is applied, the weakest fiber breaks and the load carried by that fiber is shared equally by all the remaining fibers, which can trigger further failures. Through gradual increase of the load, therefore, the model goes through intermittent stable states, which are subsequently perturbed by increasing the load slowly. In going from one stable state to another, the number of fibers that break is the avalanche size (s). The size distribution of this follows a power-law statistics $P(s) \sim s^{-5/2}$ for $s \rightarrow \infty$ [13]. The avalanche dynamics continues until the load per fiber value $\sigma = W/N$ reaches a critical limit σ_c , when the entire system breaks down.

The avalanches are arranged in the ascending order to estimate h -index values for the stress level σ below σ_c . It was shown ([9]) that the terminal value of $h (= h_f)$ at the critical point ($\sigma = \sigma_c$) follows a scaling relation $h_f \sim \sqrt{N}/\log(N)$. Here, we look for the scaling of h with the critical interval $(\sigma_c - \sigma)$. It turns out that the scaling relation (Eq. 1a) fits very well with our numerical data (see Figure 2). We also attempted the

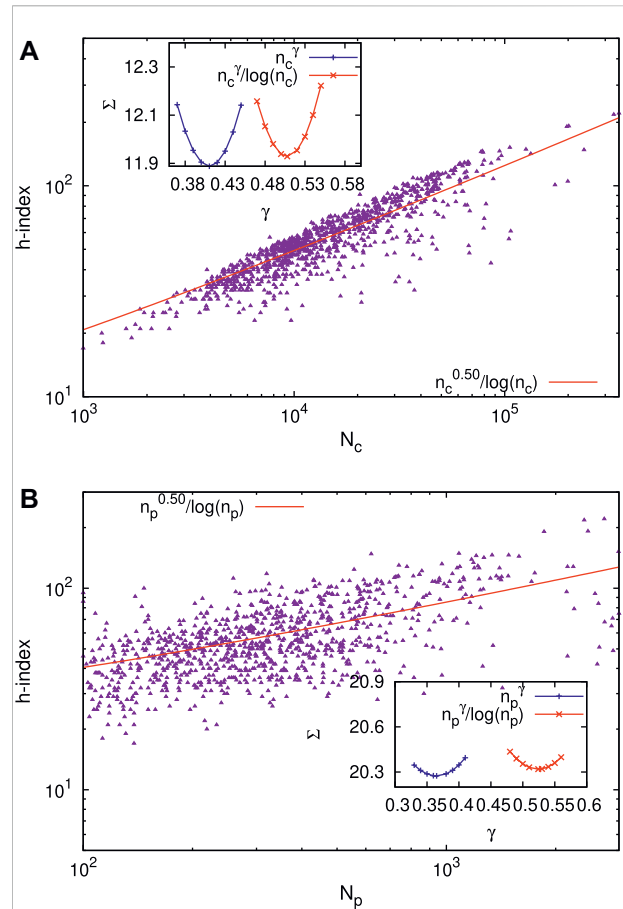


FIGURE 8
Scaling behavior for the h -index (in the range $17 \leq h \leq 221$) of 1000 scientists with the total number of citations N_c (in the range $996 \leq N_c \leq 348680$) in (A) and with the total number of papers N_p (in the range $100 \leq N_p \leq 2987$) in (B). The data are taken from Google Scholar in June 2021. The figures show the best fits to $h \sim \sqrt{N_c}/\log N_c$ in (A) and to $h \sim \sqrt{N_p}/\log N_p$ in (B). The insets show the statistical deviations $\Sigma = \sqrt{\sum_i (h_d(i) - h_\gamma(i))^2 / \sum_i 1}$ to the scaling forms $h \sim N_c^\gamma$ (or $h \sim N_p^\gamma$) and $h \sim N_c^\gamma / \log(N_c)$ (or $h \sim N_p^\gamma / \log(N_p)$) with respect to γ , where i represents the individual scientist with the corresponding $h (= h_d)$ value obtained from Google Scholar and the scaling fit (h_γ) from the aforementioned scaling relations. The inset shows that although the statistical deviation does not become minimum for the scaling with the log correction term (presumably due to inclusion of some tail-end points), the best fit for γ assumes the desired value $1/2$.

scaling fit without the $\log(N)$ term in the aforementioned relation (see inset of Figure 2), and it clearly worsens the data collapse.

Figure 3 shows the scaling behavior for the average h -index (in the range $4 \leq h \leq 20$) at the breaking point ($\sigma = \sigma_c$) of the bundles with the total number N (in the range $200 \leq N \leq 20000$) of fibers in the equal-load-sharing FBM (with uniform distribution of fiber-breaking thresholds) considered here (cf [9]). The figure shows the best fit of h to $\sqrt{N}/\log N$.

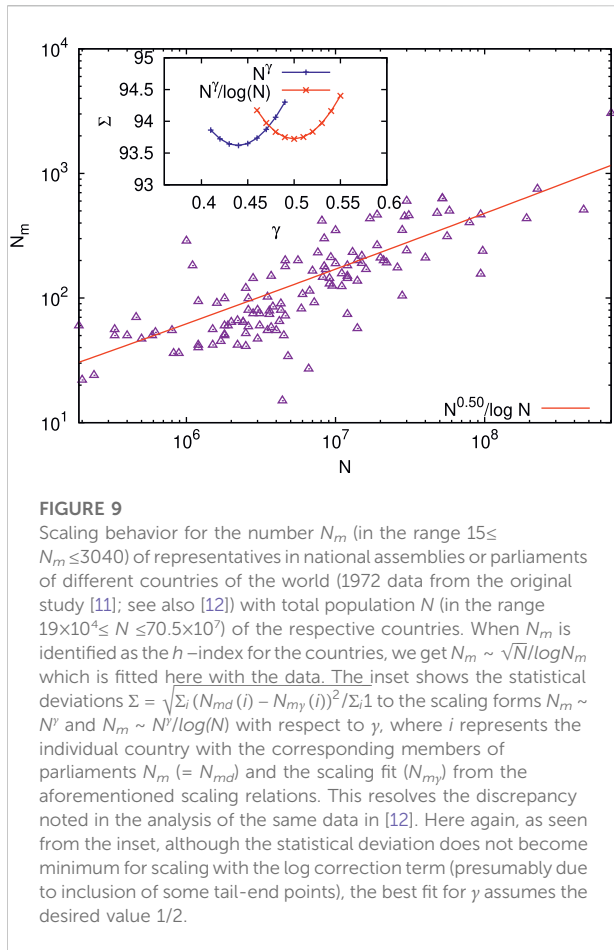


FIGURE 9

Scaling behavior for the number N_m (in the range $15 \leq N_m \leq 3040$) of representatives in national assemblies or parliaments of different countries of the world (1972 data from the original study [11]; see also [12]) with total population N (in the range $19 \times 10^4 \leq N \leq 70.5 \times 10^7$) of the respective countries. When N_m is identified as the h -index for the countries, we get $N_m \sim \sqrt{N}/\log N_m$ which is fitted here with the data. The inset shows the statistical deviations $\Sigma = \sqrt{\sum_i (N_{md}(i) - N_{my}(i))^2 / \sum_i 1}$ to the scaling forms $N_m \sim N^\gamma$ and $N_m \sim N^\gamma / \log(N)$ with respect to γ , where i represents the individual country with the corresponding members of parliaments $N_m (= N_{md})$ and the scaling fit (N_{my}) from the aforementioned scaling relations. This resolves the discrepancy noted in the analysis of the same data in [12]. Here again, as seen from the inset, although the statistical deviation does not become minimum for scaling with the log correction term (presumably due to inclusion of some tail-end points), the best fit for γ assumes the desired value $1/2$.

2.2 Percolation clusters in the 2D lattice

Here, we consider two-dimensional (2D) site percolation in a square lattice with site occupancy probability p ($0 < p < 1$). For a given p , we measure the cluster distribution, and hence, the h -index is computed (here, h -index measures h number of clusters each having cluster sizes greater than equal to h). In our simulation, we took four different system sizes ($N = 4000 \times 4000$, 8000×8000 and 16000×16000), and the h -index was estimated for different p values.

Here, we carry out the finite size scaling analysis [14] (see also [15]) of the h -index for different system sizes. The critical exponent α is determined by the fit to scaling relation (Eq. 1b) for all the data points of different sizes ($N = 4000 \times 4000$, 8000×8000 and 16000×16000). We obtain a good data collapse with $\alpha = 3$ and different scaling functions f for $p > p_c$ and $p < p_c$ while they converge to the same value at $p = p_c$ (see Figure 4). The inset of Figure 4 shows the scaling fit to relation (Eq. 1b) without the $\log(N)$ term, and it is observed that the right hand side scaling fit is not that well.

The cluster size distribution in the percolation problem has also been studied for estimating h -index scaling with the total number N of lattice sites at the percolation threshold of site percolation on the square lattice. We, of course, find here the best fit scaling form to be $h \sim \sqrt{N}/\log N$ (see Figure 5).

2.3 Paper citations

We first analyze the data for the h -index and its scaling with the total number of publications N_p and of citations N_c for the 100 scientists (in mathematics, physics, chemistry, medicine, biology, economics, and sociology, including those of 20 Nobel laureates in those subjects) given in [5]. Next, we collected (from May to June 2021) the same kind of data¹ for 1000 scientists (mostly physicists) in all the aforementioned subjects from Google Scholar.

Figures 6, 7 show the scaling behavior for the h -index (in the range $20 \leq h \leq 222$) of the 100 scientists (data taken from [5]), with the total number of citations N_c (in the range $1819 \leq N_c \leq 323473$). Our analysis shows the best fit to $h \sim \sqrt{N_c}/\log N_c$.

Figures 8A, B show the scaling behavior for the h -index (in the range $17 \leq h \leq 221$) of 1000 scientists with the total number of citations N_c (in the range $996 \leq N_c \leq 348680$) in Figure 8A and with the total number of papers N_p (in the range $100 \leq N_p \leq 2987$) in Figure 8B. The data are taken from Google Scholar in June 2021. The figures show the best fits to $h \sim \sqrt{N_c}/\log N_c$ in Figure 8A and to $h \sim \sqrt{N_p}/\log N_p$ in Figure 8B.

2.4 Number of representatives in the national assemblies

Finally, we note that if the number N_m of representatives in the national assemblies or in the parliaments of different countries of the world is identified as the h -index for the respective country, having population N , then we find (see Figure 9) that the data analyzed in [11]² show the best fit of N_m to $\sqrt{N}/\log N$. This also resolves the discrepancy noted in the analysis of the same data ([12] and references therein).

¹ The data will be available on request to corresponding author.

² The data extracted from this paper can be found in <https://sciencehistory.epfl.ch/physics-and-sociology/>

TABLE 1 Values of the pre-factors a and \bar{a} for the h -index scaling fits.

	FBM avalanche size	Percolation cluster size	Citation size	Parliament size
Fit to $h = a\sqrt{N}$ $a =$	0.15 ± 0.10	0.15 ± 0.10	0.45 ± 0.15	0.10 ± 0.08
Fit to $h = \bar{a}\sqrt{N}/\log(N)$ $\bar{a} =$	2.00 ± 1.00	2.50 ± 1.00	3.50 ± 0.50	2.00 ± 1.25

3 Summary and conclusion

We have studied here the scaling behaviors of the Hirsch index h for inequalities [1] applied to the unequal distributions of responses and statistics in different physical (fracture and percolation) and social systems (citations and parliament sizes). We have shown that the Hirsch index follows remarkable off-critical Widom–Stauffer scaling in these systems that are widely different in terms of their dimensionality and symmetry.

We have studied avalanche or cluster sizes in physical systems like the fiber bundle models (see e.g. [2, 10]) and percolating systems (see e.g. [3, 14]). Indeed, surprising successes of such Hirsch-like social inequality measures were already seen (see e.g. [9, 15]) in predicting the global failures in fiber bundles and the self-organized critical points of sand-pile systems. We show, in this study, from the Monte Carlo simulation results (see Section 2.1 and Section 2.2) that the Widom–Stauffer-like scaling relation (Eq. 1a,b) of the Hirsch index h (defined here for the size s distributions n_s , through $h = s = n_s$) fits remarkably well with the system size N (see Figures 2, 4). We also show (see the insets of Figures 2, 4) that the scaling collapse breaks down without the $\log N$ term in (Eq. 1) (visibly clear for the fiber bundle results in Figure 2).

As it is well known, the Hirsch index (h) was introduced originally to measure the inequalities in success (through citations in the subsequent literature) of the contributions (papers) of individual scientists. The citation function (see Figure 1), is a well-documented non-linear and monotonically decaying function (cf. Zipf law [16]). h corresponds to the fixed point of this citation function. The analytical study [4] and numerical data analysis (see e.g. [5]) suggested $h \sim \sqrt{N_c} \sim \sqrt{N_p}$ for any author having N_p total papers and N_c total citations. Since for different scientists, we do not have any knowledge about the critical intervals (like $|\sigma_c - \sigma|$ or $|p_c - p|$ in relations (Eq. 1)); here, the detailed data for h are fitted to the relations (Eq. 2) (see Section 2.3, Figures 7, 8A,B), assuming the critical interval to be zero for the chosen authors. We show that a better fit is $h \sim \sqrt{N_c}/\log N_c \sim \sqrt{N_p}/\log N_p$ (see relations (Eqs. 2a,b)). It may be noted that analysis of a larger data set (see e.g. [17]) indicated a significantly lower value (≈ 0.42) of γ . As may be seen from our analysis (see Figures 7, 8), this is probably due to the choices of data corresponding to very large deviations from the respective critical points to fit the power law scaling form and the missing $\log N$ term (in relations (Eq. 2)) in the scaling form. It suggests, in absence of this information, extending the data sets for scientists who may not

be that competitive may lead to wrong conclusions. Although all the data for the h -index of FBM avalanches and percolation clusters fit so well with the Widom–Stauffer scaling (Eq. 2) with $\gamma = 1/2$, when the data for their critical points are fitted to $h \sim N^\gamma$ (see insets of Figures 3, 5), one gets best fits for γ around 0.40 and 0.48, respectively.

We also show that when the number N_m of the members of national assemblies or of parliaments for different countries of the world is identified effectively as their Hirsch index h , then N_m indeed would scale (see Figure 9 in Section 2.4) with the total population N as $\sqrt{N}/\log N$ for the respective countries. This observation should help comprehending the discrepancies with the proposed $N_m \sim \sqrt{N}$ relationship, reported in a very recent analysis [12].

For an additional check for the $\log N$ term in the scaling behavior of the Hirsch index, we fitted the data for h in fiber bundle models (Section 2.1, for stress values σ near the failure point σ_c), for percolation systems (Section 2.2, for site occupation concentrations p both above and below the percolation threshold p_c), paper citation data (Section 2.3, for authors with Kolkata index values k both above and below the critical value k_c), and parliament membership data (Section 2.4) to the scaling forms $h = a\sqrt{N}$ and $h = \bar{a}\sqrt{N}/\log(N)$, where N denotes the appropriate system size (total number N of fibers in the FBM, total number N of lattice sites, total number of citations N_c of the author, or the total population N of the country). The estimated best fit values of the pre-factors a or \bar{a} are given in Table 1. The range of error bars in these estimates of a and \bar{a} is such that more than 95% of the data points fall within the indicated ranges. If we assume the value of the pre-factor to be the same across the systems considered (Young's asymptotic result [4] suggested the value of the pre-factor a for citation to be about 0.54), the observed fitting error ranges for other systems (see Table 1) then clearly suggest the fit with the $\log N$ term to be more appropriate. The average value of the pre-factor ($\bar{a} \approx 3.0$ in Table 1) fits reasonably across the systems considered and also agrees with the values of $f(0)$ in Figure 2 (for FBMs), Figure 4 (for percolating systems), and Figure 6 (for paper citations by an individual scientist). The comparative agreements of the values of a (with large percentile errors) and \bar{a} ($= f(0)$ in Figures 2, 4, 6 having smaller percentile errors) in Table 1 clearly indicate that the fitting of the h -index to the scaling relation (Eq. 1) is much better. These errors due to the fittings with scaling relations (Eq. 2), therefore, have become much more prominent for

parliament member number data, where the critical intervals or the country's distance from the respective critical point are completely unknown.

In conclusion, we have explored the best fit of the Hirsch index values h for system sizes N (or N_c) with a Widom–Stauffer like finite size scaling form (Eq. 1). This is essentially based on the Monte Carlo simulation data collapse in the fiber bundle model and percolating systems (for σ or p away from σ_c or p_c , respectively; see Figures 2, 4) and data analysis for citations of 100 scientists (k away from k_c ; see Figure 6) analyzed in [5]. Data for parliament member numbers (identified as the corresponding country's h index; see Figure 9) are fitted to the relation (Eq. 2) as the equivalent critical interval ($[\sigma_c - \sigma]$, $|p_c - p|$, or $|k_c - k|$) is unknown. We find the scaling fit (to relations Eqs 1, 2) deteriorate considerably if the $\log N$ term is dropped. We give in Table 1 the estimated error in the pre-factors a and \bar{a} , and assuming the pre-factor to have the same value across the systems considered here, we again find the finite size scaling relation for the Hirsch index h with the $\log N$ term ($h = \bar{a}\sqrt{N}/\log N$ and $\bar{a} \equiv f(0)$ of the Widom–Stauffer relation Eq. 1 for finite system size N) to be more appropriate for the scaling behavior of the Hirsch index near the respective critical points.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials; further inquiries can be directed to the corresponding author.

References

- Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* (2005) 102(46):16569–72. doi:10.1073/pnas.0507655102
- Biswas S, Ray P, Chakrabarti BK. *Statistical physics of fracture, breakdown, and earthquake: Effects of disorder and heterogeneity*. New York, NY, USA: John Wiley & Sons (2015).
- Stauffer D, Aharony A. *Introduction to percolation theory*. Boca Raton, FL, USA: CRC Press (2018).
- Yong A. A critique of hirsch's citation index: A combinatorial fermi problem. *Not Am Math Soc* (2014) 61(9):1040–50. doi:10.1090/noti1164
- Ghosh A, Chakrabarti BK. Limiting value of the Kolkata Index for social inequality and a possible social constant. *Physica A: Stat Mech its Appl* (2021) 573:125944. doi:10.1016/j.physa.2021.125944
- Banerjee S, Chakrabarti BK, Mitra M, Mutuswami S. Inequality measures: The Kolkata index in comparison with other measures. *Front Phys* (2020) 8:562182. doi:10.3389/fphy.2020.562182
- Dunbar RIM. Neocortex size as a constraint on group size in primates. *J Hum Evol* (1992) 22(6):469–93. doi:10.1016/0047-2484(92)90081-j
- Bhattacharya K, Ghosh A, Monsivais D, Dunbar RIM, Kaski K. Sex differences in social focus across the life cycle in humans. *R Soc Open Sci* (2016) 3(4):160097. doi:10.1098/rsos.160097
- Biswas S, Chakrabarti BK. Social inequality analysis of Fiber Bundle Model statistics and prediction of materials failure. *Phys Rev E* (2021) 104(4):044308. doi:10.1103/physreve.104.044308
- Pradhan S, Hansen A, Chakrabarti BK. Failure processes in elastic fiber bundles. *Rev Mod Phys* (2010) 82(1):499–555. doi:10.1103/revmodphys.82.499
- Taagepera R. The size of national assemblies. *Soc Sci Res* (1972) 1(4):385–401. doi:10.1016/0049-089x(72)90084-1
- Margaritondo G. Size of national assemblies: The classic derivation of the cube-root law is conceptually flawed. *Front Phys* (2021) 8:606. doi:10.3389/fphy.2020.614596
- Hemmer PC, Hansen A. The distribution of simultaneous fiber failures in Fiber Bundles. *J Appl Mech* (1992) 59:909–14. doi:10.1115/1.2894060
- Stauffer D. Scaling theory of percolation clusters. *Phys Rep* (1979) 54(1):1–74. doi:10.1016/0370-1573(79)90060-7
- Manna SS, Biswas S, Chakrabarti BK. Near universal values of social inequality indices in self-organized critical models. *Physica A: Stat Mech its Appl* (2022) 596:127121. doi:10.1016/j.physa.2022.127121
- Zipf GK. *Selected studies of the principle of relative frequency in language*. Cambridge, MA: Harvard University Press (1932).
- Radicchi F, Castellano C. Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics* (2013) 97(3):627–37. doi:10.1007/s11192-013-1027-3

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors are grateful to Parongama Sen for useful comments and suggestions. BC is thankful to the Indian National Science Academy for their Senior Scientist Research Grant.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.