JOURNAL COMPUTER SCIENCES INSTITUTE

JCSI 23 (2022) 152–164
Received: 13 April 2022
Accepted: 10 May 2022

Comparative study of scaling parameters and research output of selected highly- and moderately-cited individual authors

Badanie porównawczych parametrów skalowania i dorobku badawczego wybranych wysoko i umiarkowanie cytowanych autorów

Keshra Sangwal*
Department of Applied Physics, Lublin University of Technology, ul. Nadbystrzycka 38, 20-618 Lublin, Poland

Abstract

The real data of cumulative citations l_n of selected nth paper of individual N papers published by some highly- and moderately-cited individual authors are analyzed to compare Hirsch and Hirsch-type indices h, h_1, h_f and h_m, and citation radii R and R_f from consideration of: (1) the number A_n of coauthors of the paper, (2) the normalization of citations l_n and cumulative fraction l_{nf} of citation of the nth paper by mean and median citations of the citations l_n of all N_c cited papers, and (3) the determination of effective rank n_{eff} of the l_{nf} citations. Analysis of the $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$ data was also carried out by using a Langmuir-type function $l = l_0[1-\alpha Kn/(1+Kn)]$, where l denotes the citations l_n and l_{nf} of all cited N_c papers arranged in the decreasing order, α is an effectiveness parameter, K is the so-called Langmuir constant, n denotes the rank n or n_{eff} of citations and l_0 is the value of l when n or n_{eff} approaches zero. For a comparison of the publication output of different authors it was found that the h_m index is more consistent than other indices, and it can be normalized to account for the publication career of different authors. However, Langmuir-type function is not adequate for comparison of the publication output of different authors because it describes the rank-order distribution patterns satisfactorily in terms of two parameters. To compare the publication output of different authors independent of their career length t, it is suggested to use scaling parameters h/t, h_f/t and h_m/t.

Keywords: Citation analysis; Citation distribution; Coauthorship; h-type indices; Langmuir-type function

Streszczenie

Przeanalizowano dane liczby cytowań kumulacyjnych l_n wybranego n-tego artykułu spośród N indywidualnych artykułów opublikowanych przez niektórych wysoko i umiarkowanie cytowanych pojedynczych autorów. Do porównania użyto wskaźników Hirscha h i Hirscha-podobnych h_1, h_f i h_m, oraz promienia cytowań R i R_f pod względem: 1) liczby A_n współautorów artykułów, 2) normalizacji liczby cytowań l_n i ułamka kumulacyjnego l_{nf} cytowania n-tego artykułu średnimi i środkowymi cytowaniami l_n wszystkich cytowań N_c artykułów, 3) określenia efektywnego rzędu n_{eff} cytowań l_{nf}. Analiza danych $l_n(n)$, $l_{nf}(n)$ i $l_{nf}(n_{eff})$ była prowadzona również wedle funkcji typu Langmuira $l = l_0[1-\alpha Kn/(1+Kn)]$, gdzie: l oznacza l_n i l_{nf} cytowania wszystkich cytowanych N_c artykułów umieszczonych w malejącym porządku, α parametr efektywności, K to tak zwana stała Langmuira, n oznacza rząd n lub n_{eff} cytowań oraz l_0 to wartość l, gdy n lub n_{eff} dąży do zera. Do porównania dorobku publikacyjnego różnych autorów stwierdzono, że wskaźnik h_m jest bardziej miarodajny od pozostałych oraz, że można je znormalizować w celu uwzględnienia kariery publikacyjnej różnych autorów. Funkcja typu Langmuira nie jest właściwa jednak do porównania dorobku publikacyjnego różnych autorów ponieważ opisuje ona w sposób zadowalający rozkład kolejności rzędu ich artykułów przy pomocy dwóch parametrów. Do porównania dorobku publikacyjnego różnych autorów, niezależnego od długości ich kariery t, zaproponowano stosowanie parametrów skalowania h/t, h_f/t and h_m/t.

Słowa kluczowe: Analiza cytowań; Rozkład cytowań; Współautorstwo; Współczynniki typu h; Funkcja typu Langmuira

*Corresponding author
Email address: **k.sangwal@pollub.pl** (K. Sangwal)

1. Introduction

For promotion/recruitment of faculty/research positions and award of research grants it is desired to compare the research output of candidates working in a scientific field. A commonly used measure for this purpose is the h index, proposed by Hirsch [1], which is defined as the number of papers of nth rank with citations $l_n \geq h$. A convenient way to determine the h index of an author is to look for the value of the rank n of the paper when $n \leq l_n$ from the plot of the decreasing number l_n of citations received by the nth paper against all of his/her N_c cited papers (rank–size distribution plots). The main advantages of the h index are that it is a single number characterizing publications and citations of an author and it is insensitive to uncited or relatively poorly cited papers. However, the citation distribution of an author is frequently skewed either with a few highly cited papers or a large number of papers with few citations. Consequently, since an n-ranked paper enters in the counting of the h index of an author, citations outside the h-core (i.e. citations $l_n > h$ received by papers $n < h$ and $l_n < h$ received by papers $n > h \geq N_c$) are not used anywhere.

In order to overcome the above disadvantage of the h index its several variants, simultaneously retaining its advantages, have been widely proposed and discussed (for example, see: refs. [2-13]).

It is well known that the Hirsch index h of an author increases with his/her publication duration t [1,5,14-17], typical h indices of researchers publishing in different research fields are different [1], and researchers publishing in large collaborations usually have high values of their h index [1,18,19]. In fact, these factors make the comparison of scientific research output of authors of different academic career length working across various research disciplines difficult. In order to compare the research output of researchers publishing a constant number of papers per year of similar quality during their publication duration t, Hirsch [1] proposed the parameter $m = h/t$ as a useful measure. From an analysis of the behavior of h index as a function of academic age t of about 1400 Italian physicists, Mannela and Rossi [16] found that a time-scaled index $h/t^{1/2}$ is related to the h index.

All citation-related measures, including the h index and its different variants, for the publication output of an author assume that all papers receiving citations are written by him/her alone. However, in the case of multiple-authored papers, it is unfair to award full credit to each author because this method in reality penalizes authors who publish alone. Therefore, devising of a fair method of counting of contributions of individual authors in multi-authored papers has drawn considerable attention for over four decades (for example, see: refs. [18,20-26]). The problem is also complicated because single-authored papers published in different journals with high impact factors usually earn the lowest number of citations and the number L of cumulative citations increases with increasing number A of coauthors, following approximately the relation: $L = L_0(0.2A)^{1/3}$ [27]. In this relation L_0 is a normalizing factor related to the journal.

The culture of authorship of papers in different disciplines is not the same. The following authorship patterns are usually observed [24,28]: (1) junior researchers are the first authors and group leaders are the last authors, (2) senior researchers are the first authors followed by junior researchers, (3) authors in multi-authored papers are listed in the order of decreasing contribution of coauthors involved their publication, and (4) all authors are arranged alphabetically, especially in large collaborations as in nuclear physics experiments. In case (2) it is not always possible to establish the contributions of the first, senior, corresponding authors. However, in general, contributions of different authors in multi-authored papers frequently remain unknown and the information that one usually has to determine the credit of authors in different multi-authored papers is the authors' list in the papers. Therefore, assigning due credit to the coauthors of papers in different disciplines is a problem in citation analysis.

For publications involving a large number of coauthors, Hirsch [1] suggested to normalize individual authors to normalize their h indices by a factor reflecting the average number of authors. Counting of contributions according to the order of the authors has also been proposed [21,27,29-33]. In this approach the contribution of the first author is always dominant in a multi-authored paper but equal contributions of the first and last (corresponding) author can also be considered [31].

It is well known that the average number of citations per paper differs among various scientific disciplines due to their citation behavior (for example, see: refs. [1,4,34-38]). Using the total number of citations [35,36] or distributions of citations to the papers published in different fields [4,34,37-39] different scaling parameters have been proposed to compare the research output of researchers in various fields. The average number of citations in some studies [1,4,34,35,37,38] whereas median or geometric mean of citations in others [36] has also been proposed as an effective scaling parameter in different fields.

To account for the effect of multiple coauthorship through the h index, Hirsch [40] proposed $\hbar$ (hbar) index as the number of an individual author's papers that have citations greater than or equal to the $\hbar$ index of all coauthors of each paper. However, this approach considers papers instead of authors. Following the idea advanced by Hirsch [1] of normalizing individual authors to normalize their h indices by a factor reflecting the average number of authors, Batista et al. [18] proposed to divide the h index of an individual author by the number of authors of the paper in the h-core. The resulting so-called h_1 index is defined as: $h_1 = h^2/A_h$, where A_h is the total number of coauthors of a particular author in the h-core. The main problem with this h_1 index is that its value is enormously changed for authors with some papers with a large number of coauthors. Egghe [41] proposed to count the contribution of multiple authors considering either fractions l_{nf} of citations (i.e. $l_{nf} = l_n/A_n$, where A_n is the number of authors of a paper) or fractions n_{eff} of rankings of all n-ranked papers (i.e. $n_{eff} = n/A_n$). Arrangement of the fractions l_{nf} of citations to all n-ranked papers of an author in decreasing order gives his/her new Hirsch-type index, namely the h_f index, which is again an integer as the h index. Similarly, counting of the rank n of each paper fractionally as inverse of the number A_n of its authors results in the so-called h_m index, which is a noninteger number [42,43].

Performance of multi-authorship indices for the ranking of authors has been discussed in the literature [26,44-46]. Abramo et al. [44] recommended to use indicators or evaluation methods that take into consideration authors' contributions. Sahoo [26] proposed I-index from consideration of an author's percentage share in the total citations received by his/her papers. He observed that equi-distribution of credit among the coauthors of a paper would give the most probable value of his I-index (with an associated small standard deviation which decreases with increasing h-index). Dunaiski et al. [45] reported best results with equal contribution of co-authors to a paper's score, independent of impact indicator used to compute paper scores.

The h index of authors is an empirical measure of their publication output. Since the h index of an author is defined in terms of the highly cited papers, its prediction is essentially associated with the description of the rank-order distribution of citations of his/her papers. This approach has been used by Egghe [47], and Egghe and Rousseau [48,49], who used Lotka's law for the rank-order distribution of citations. However, Sangwal [50] observed that Lotka-type function poorly describes real rank-order distributions of citations of different authors. From an examination of real citation distributions of authors Burrell [51] also arrived at a similar conclusion.

Following the concepts of processes of adsorption of additives during crystal growth, Sangwal [50] proposed four novel functions to describe rank-order distributions of citations of different authors. He observed that, among these functions, a Langmuir-type function satisfactorily describes the $l_n(n)$ data for different authors and that the Langmuir constant K of this function characterizes the citation behavior of an author. This function was later employed to explain rank-order distributions of journals published in different countries [52] and rank-order distributions of cumulative citations l_n and fractions l_{nf} of citations of papers recived by multi-authored papers published by moderately-cited authors working in different scientific fields [53]. Since the h, h_f and h_m indices of an author are related to the three distributions $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$, respectively, constructed from the bibliometric data of his/her publication output, it is interesting to analyze the research output of some highly- and moderately-cited authors using the proposed Langmuir-type function.

The aim of the present paper is fourfold: (1) to verify whether normalization of distribution of cumulative citations l of individual papers with total citations L and the cumulative fractions L_f of citations of the papers of an author, based on equal contribution of different authors by their mean and median citations, yields a universal citation distribution, (2) to analyze the distribution of normalized citations l and l_f using Langmuir-type function, (3) to compare the research output of selected authors using different citation indices proposed in the literature, and (4) to propose an objective measure for characterization (and comparison) of the publication output of different authors from distribution of cumulative citations l_n of his/her N_c cited papers published during their entire publication career.

2. The Langmuir-type function

Langmuir-type function of rank-order distribution of items is based on the concepts of adsorption processes involved during crystal growth. The basic concepts used in the derivation of this function have been described earlier [50,52,53]. In the context of citation distribution the Langmuir function is based on the following postulates:

(1) The N papers of an author have the same number s_{max} of possible active citation sites, and these possible sites for the papers can be represented in the form of a two-dimensional set of successive papers arranged at equal distance along the x axis such that s_{max} equi-spaced points, denoted as sites l_n, stacked along the y axis represent citations received (filled circles) and unreceived (open circles) by each of the N papers. Figure 1 schematically illustrates this arrangement of N papers in the form of decreasing citations

(2) The maximum number of cumulative l_0 citations belongs to the paper of rank $n = 0$ represented by s_{max} possible active sites, no citations are received by the papers of rank $n > N_c$ represented by s_{ad} sites, whereas $0 < l_n < l_0$ citations are received by the papers $0 < n < N_c$ represented by (s_{max}–s_{ad}) sites. If we define the ratio of the number of s_{ad} (covered) sites to the total number s_{max} of possible active sites as the coverage θ of active sites (i.e. $\theta = s_{ad}/s_{max}$), the coverage θ may be considered as a measure of unreceived citations by the nth paper.

(3) The rank n of a paper is a measure of probability of receiving citations by the paper. This assumption means that the number of l_n citations received by the nth rank paper decreases with increasing value of n whereas l_0 citations are produced by the paper of rank $n = 0$.

(4) The process of unreceiving (inhibition or blocking) of citations may be described by the usual isotherms applied in adsorption of adsorbant atoms, ions or molecules on active adsorption sites on the surface of a solid.

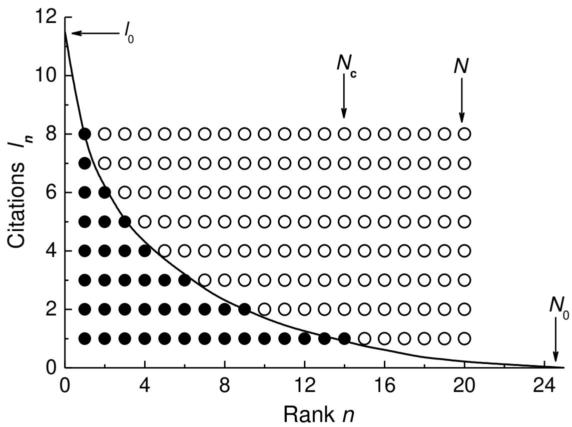

Figure 1: Graphical illustration of the number l_n of citations received by a set of n papers. Filled and open circles denote received (blocked) and unreceived (uncovered) citations, respectively. Dependence of l_n citations on rank n of papers is shown by the curve starting from citations l_0 at rank $n = 0$ and terminating at rank $n = N_c$. Note that N_c is the maximum number of papers that received citations ($N_c = 14$ here).

With the above assumptions the relationship between the number l of citations received by a paper and the coverage θ of blocked citation sites may be given by

$$l = l_0(1 - \alpha\theta) , \tag{1}$$

where l_0 is the maximum number of citations received when $\theta = 0$ and α is an effectiveness parameter for unblocking of citation sites (see below). When the coverage θ of citation sites follows Langmuir adsorption isotherm

$$\theta = \frac{Kn}{1+Kn}, \tag{2}$$

one obtains from Eq. (1) the usual Langmuir-type (LT) function relating the number l of citations with the rank n of the paper in the form

$$l = l_0\left[1-\alpha\left(\frac{Kn}{1+Kn}\right)\right]. \tag{3}$$

In Eqs. (2) and (3) the parameter K, usually known as the Langmuir constant, characterizes the behavior of citations generated by the papers whereas n is the rank of the paper. The units of K are inverse of paper rank (i.e. paper-rank^{-1}) and are determined by the way the paper rank n is expressed. If one defines a dimensionless paper rank $x = n/N_c$, Eq. (3) may be expressed as

$$l = l_0\left[1-\alpha\left(\frac{K'x}{1+K'x}\right)\right], \tag{4}$$

where the new dimensionless Langmuir constant $K' = KN_c$, with N_c as the number of papers which receive citations ($N_c < N$). The dimensionless Langmuir constant K' for citations of different paper–citation systems is related to their corresponding dimensionless differential energy Q by

$$Q = \ln K' = \ln(KN_c). \tag{5}$$

The effectiveness parameter α may be given by (cf. Eq. (1))

$$\alpha = \frac{1}{\theta}\left(1-\frac{l_i}{l_0}\right), \tag{6}$$

where l_i is the number of unreceived citations by a paper. In Eq. (6) when the active citation sites are blocked completely, i.e. $l_i = 0$, depending on the value of coverage θ the following situations are possible regarding the values of α: (1) when the coverge $0 \le \theta \le 1$, $\infty \ge \alpha \ge 1$, and (2) when $\theta \ge 1$, $0 \le \alpha \le 1$. In the latter case, the highest value of unity for α is achieved when $\theta = 1$ whereas $\alpha \to 0$ when $\theta \to \infty$.

When $l/l_0 = 0$, Eq. (4) gives

$$K' = KN_c = (\alpha-1)^{-1}. \tag{7}$$

According to relation (7) the dimensionless Langmuir constant $K' = KN_c$ is inversely proportional to $(\alpha-1)$ with the proportionality constant $(KN_c)_0 = 1$ such that $\ln(KN_c)_0 = 0$. Since K and K' are always positive, Eq. (7) implies that $\alpha \ge 1$. This inference is consistent with our earlier findings [50,52,53].

3. Bibliometric data and their analysis

For the analysis we used examples of bibliometric data on the cumulative citations received by papers published by 11 selected authors. Of these, six authors are from those who published first single-authored original papers which received more than 20000 citations each up to 2013. These authors were selected from the authors of first single-authors of top-cited papers, which showed mainly an initial increase and then, after going through a maximum value, a slow decrease in the number of their yearly citations in successive years [54]. The other five authors are from those who were nominated professors in chemistry by the President of Poland in 2013 and received citations lower than 2000 each up to 2012 [53]. Hereafter these two groups of authors are referred to as highly- and moderately-cited authors, respectively.

The Thomson Reuters' Web of Knowledge database was employed to collect the basic bibliometric data. The data represent papers published till 2013 and 2012 by the highly- and moderately-cited authors, respectively. The data were collected during 21-24 October 2014 and 1-15 April 2014, respectively. An author's authorship of each paper was identified unambiguously following his/her affiliation, research area and names of frequent coauthors. Despite the fact that these bibliometric data were collected some years ago, more recent data are not included in the analysis in view of long publication careers, lying between 32 and 68 years, of most of the authors (see Table 1).

The basic bibliometric data for an author comprised the number N of cumulative papers, the number L of cumulative citations received by N_c papers (such that $N_c < N$) and the publication duration t since the publication of the first paper. In addition to the above data, the citations l_n received by individual nth paper published by an author and the number A_n of its coauthors were collected from the Web of Knowledge database. From these lists of the values of the number l_n of citations received by each of his/her nth paper and the number A_n of its coauthors, the author's cumulative citations $L = \sum l_n$, his/her cumulative fractions L_f of citations (i.e. $L_f = \Sigma l_{nf}$, where the fraction l_{nf} of l_n citations of a paper is: $l_{nf} = l_n/A_n$), and the effective rank n_{eff} of the papers (i.e. $n_{eff}(n) = \Sigma 1/A_n$, where A_n is the number of authors of the nth paper), were calculated. In view of different authorship patterns used in the lists of authors of the papers considered in this study and enormously high number of coauthors in many cases, l_{nf} and n_{eff} were calculated on the assumption of equal contributions of A_n coauthors of the nth-ranked paper fetching l_n citations. The values of the cumulative papers N published by different authors, the cumulative number N_c of their papers receiving cumulative citations L, cumulative fraction L_f of citations, and their publication duration t are given in Table 1. In columns 2 and 4 of this table N_c^a denotes the cumulative number of papers without the so-called outliers fetching exceptionally high citations whereas L^a is the cumulative number of citations received by N_c^a papers.

The real distribution of $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$ data of different papers published by the selected authors were analyzed using Eq. (3). Nonlinear least-squares fitting, involving chi-square residual, of the citation data

Table 1: Bibliometric data and various calculated indices for different authors

Author	N (N_c/N_c^a)	t	L (L^a)	L_f	h (A_h)	h_1	h_f	h_m (N_m)	R^b	R_f
Laemmli UK	93 (84/81)	45	236007 (13478)	5345.3	58 (172)	19.56	40	21.43 (55)	(65.50)	41.24
Becke AD	84 (70/67)	36	90924 (15603)	11953.3	48 (86)	26.79	40	31.42 (44)	(70.47)	61.68
Southern EM	127 (110/109)	56	38598 (6612)	3460.4	41 (120)	14.01	28	19.11 (36)	(45.88)	33.19
Felsenstein J	113 (99/94)	48	46478 (12312)	9513.5	51 (74)	35.15	30	41.92 (51)	(62.60)	55.03
Shannon RD	171 (152/150)	45	43179 (9080)	4240.0	45 (141)	14.36	28	18.30 (39)	(53.76)	36.74
Scatchard G	84 (73/72)	68	29077 (5655)	2800.6	33 (87)	12.52	25	16.03 (32)	(42.43)	29.86
MBa	103 (70/69)	15	1331 (582)	237.01	17 (572)	0.51	7	5.24 (21)	20.58 (13.61)	8.69
MCy	80 (68)	35	1117	310.06	18 (85)	3.81	9	8.06 (27)	18.85	9.93
TJe	164 (92)	15	852	390.01	18 (46)	7.04	10	8.06 (15)	16.56	11.14
BBo	80 (74)	41	963	393.35	17 (48)	6.02	10	8.34 (19)	17.51	11.19
BPa	56 (44/43)	32	688 (593)	426.33	16 (32)	8.0	11	10.32 (15)	14.80 (13.74)	11.64

[a] Values of N_c for cumulative citations L^a. [b] Values of R calculated from citations, excluding outliers as mentioned in Table 2, are given in the parentheses.

Table 2: Mean and median values of citations of different authors

Author	Citations L (with coauthors)			Citations L_f (without coauthors)	
	N range	Mean	Median	Mean	Median
Laemmli UK	≥4	166.40	87.0	65.99	9.5
Becke D	≥4	232.88	91.0	178.41	46.0
Southern EM	≥2	60.66	24.0	31.75	9.75
Felsenstein J	≥6	130.98	49.0	101.21	39.0
Shannon RD	≥6	41.66	23.0	15.98	6.0
Scatchard G	≥2	78.54	26.5	38.90	11.75
MBa	All (≥2)	22.71 (12.19)	6.0	3.39	1.5
MCy	All	16.43	6.5	4.56	2.83
TJe	All	9.17	5.0	4.24	1.75
BBo	All	13.01	7.0	5.32	3.0
BPa	All (≥2)	15.64 (13.79)	6.5	9.69	4.0

was carried out with commercially available "Origin 9.1" package. This package yields values of the fitting parameters of an equation, their standard deviations and the corresponding goodness-of-the-fit parameter R^2. In the case of Eq. (3) the best-fit parametrs, among others, are effectiveness parameters α, α_f and α_f^{ef}, and Langmuir constants K, K_f and K_f^{ef} for the $l^*(n)$, $l_f^*(n)$ and $l_f^*(n_{eff})$ data. Here l^* and l_f^* are cumulative relative citations normalized by dividing l_n and l_{nf} citations by the average (mean) or median number of citations <L> and <L_f>, respectively, obtained from the values of l_n and l_{nf} citations of N_c cited papers of an author (see Table 2). As seen from Table 2, in the case of some of the authors, especially the highly cited authors, their highly cited papers (i.e. outliers) were excluded while calculating their mean and median citations for the papers. The above normalization procedure was used to verify whether distributions of total citations l_n of nth papers of different authors converge to a single curve.

It should be mentioned that the fractions l_{nf} of citations of individual nth paper for the various authors are distinguished from those for the cumulative citations l_n of multi-authored papers by the subscript "f". Similarly, the average value <L_f> of his/her contributed citations L_f is distinguished from the average value <L> of cumulative citations L of an author by the subscript "f". However, while denoting relative citations l_n^* and l_{nf}^* the subscript n for the rank n has been omitted in different figures and tables, and the best-fit parameters of Eq. (3) for the contributed citations l_f^* of individual papers for the various professors are distinguished from those for the cumulative citations l^* of multi-authored papers by the subscript "f".

4. Results and discussion

4.1. Normalized citations of individual papers of authors and their dimensionless rank

Figures 2 and 3 show the real data of relative citations l^* and l_f^* of papers published by different highly- and moderately-cited authors, respectively, as a function of dimensionless rank n/N_c of their papers. In these figures the relative citations l^* and l_f^* of each paper represent cumulative citations l_n of each nth paper and their cumulative fractional citations l_{nf} normalized by dividing them by their mean <L> and <L_f> citations, respectively. Figure 4 shows, as an example, relative citations l^* and l_f^* of papers published by moderately-cited authors as a function of dimensionless rank n/N_c of their papers. Here the l_n and l_{nf} citations are normalized by their median <L> and <L_f> citations, respectively.

It may be seen that the $l^*(n/N_c)$ and $l_f^*(n/N_c)$ data, normalized by both mean as well as median <L> and <L_f> citations, for different authors result in broad single curves and that the spead or dispersion of the data in the y direction is relatively large in the entire n/N_c range. As indicated by the spread of the data, normalization of citations of individual papers by mean values <L> and <L_f> of citations leads to a better convergence in the entire n/N_c range for all authors (see Figures 2 and 3), but no universality of citation distributions for the authors, similar to that reported by Radicchi et al. [37] for

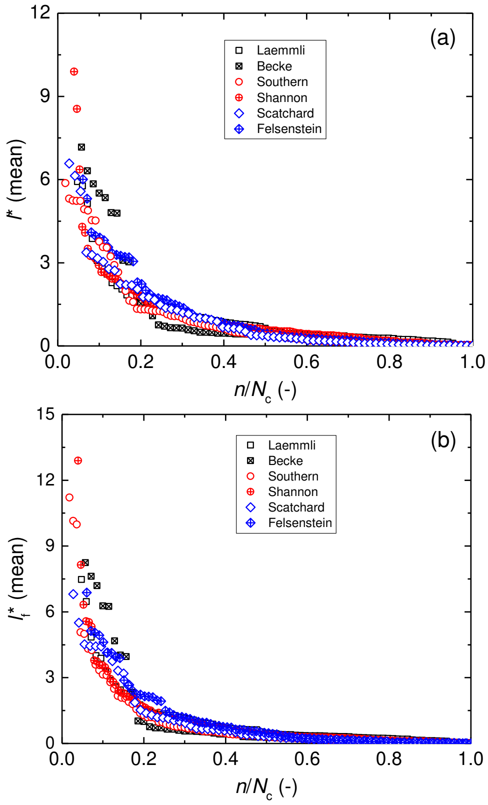

Figure 2: Normalized citations (a) l^* and (b) l_f^* of papers published by highly-cited authors as a function of dimensionless rank n/N_c of their papers. Normalization of citations was carried by the mean $<L_n>$ and $<L_{fn}>$ citations. See text for details.

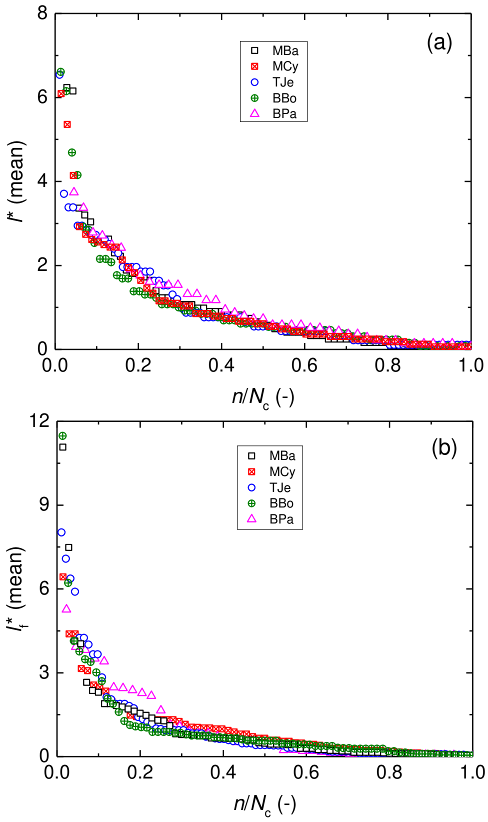

Figure 3: Normalized citations (a) l^* and (b) l_f^* of papers published by moderately-cited authors as a function of dimensionless rank n/N_c of their papers. Normalization of citations was carried by the mean $<L_n>$ and $<L_{fn}>$ citations. See text for details.

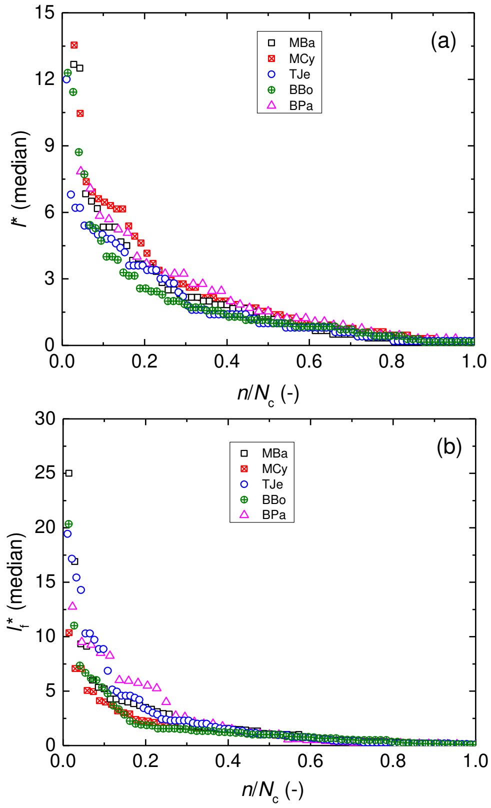

Figure 4: Normalized citations (a) l^* and (b) l_f^* of papers published by moderately-cited authors as a function of dimensionless rank n/N_c of their papers. Citations L_n and L_{fn} are normalized by their median $<L_n>$ and $<L_{fn}>$ citations.

the distribution of citations within different disciplines, is observed. However, large spread in the relative citations of different papers at low n/N_c is a general feature of these citation patterns. These dispersions are associated with different dependence of l^* and l_f^* on n/N_c for each author. This means that the distribution patterns of l^* and l_f^* are characteristic for each author, and the values of the parameters of any function such as Langmuir-type function (3) capable of describing them are representative of the author.

4.2. Rank-order distribution of normalized citations of different authors

Figures 5 and 6 show the data of relative citations l^* and l_f^* of papers, normalized by the mean values of $<L>$ and $<L_f>$ citations, of different highly- and moderately-cited authors, respectively, as a function of the rank n of their papers. Solid and dashed curves in the figures present best-fit plots for l^* and l_f^* with the values of the parameters of Eq. (3) listed in Table 3.

The distribution patterns of relative citations l^* and l_f^* normalized by the mean citations of the papers for different moderately-cited authors in Figure 6a,b are somewhat similar and frequently overlap each other. This feature may be noted from Figure 6a. However, when all papers of an author are included in the normalization of citations, the overlapping of the two citation distribution appears to improve when the first highly-cited papers are excluded from the normalization

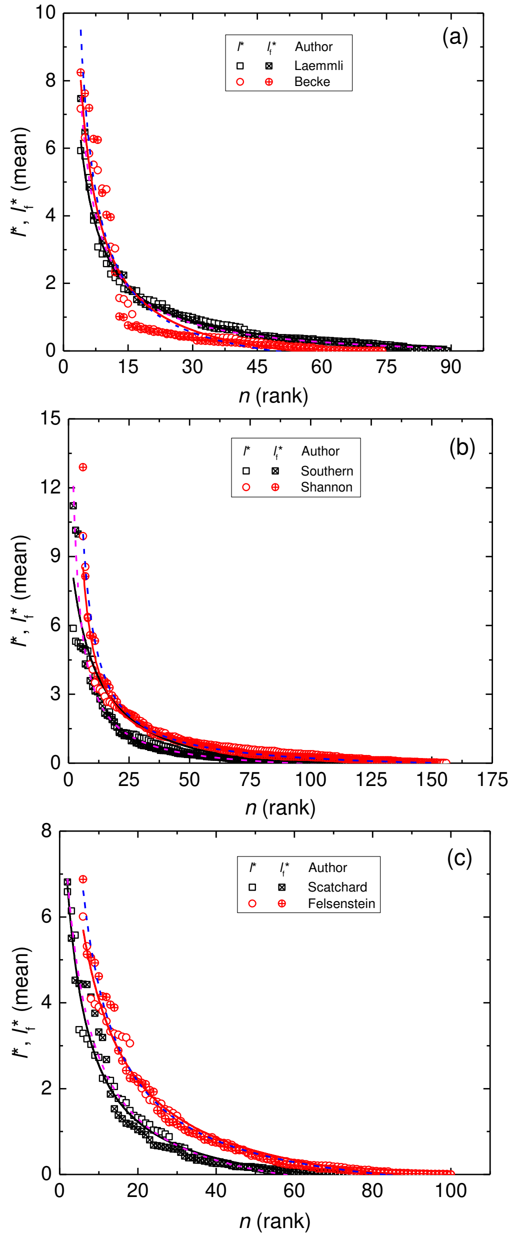

Figure 5: Plots of l^* and l_f^* of papers of different highly-cited authors against their rank n: (a) Laemmli and Becke, (b) Southern and Shannon, and (c) Scatchard and Felsenstein. Solid and dashed curves are drawn for l^* and l_f^* with the best-fit values of the parameters of Eq. (3) listed in Table 2. Cumulative citations L_n and cumulative fractionalized citations L_{fn} are normalized by mean $\langle L_n \rangle$ and $\langle L_{fn} \rangle$ citations.

(see Figure 6b).

From the trends of the distributions of relative citations l^* and l_f^* of papers published by different highly- as well as moderately-cited authors, the following inferences can be made:

(1) Mean citations of individual papers of different authors is better for comparison of their publication output.

(2) Rank-order distribution patterns of relative citations l^* and l_f^* normalized by the mean citations of the papers for different authors are different and are characterized by the two sets of three parameters: l_0^*, α and K for the $l^*(n)$ data and l_{f0}^*, α_f and K_f for the $l_f^*(n)$ data.

(3) Inclusion of citations l_n and l_{nf} of individual highly-cited papers during the analysis leads to large differences in the distributions of normalized citations.

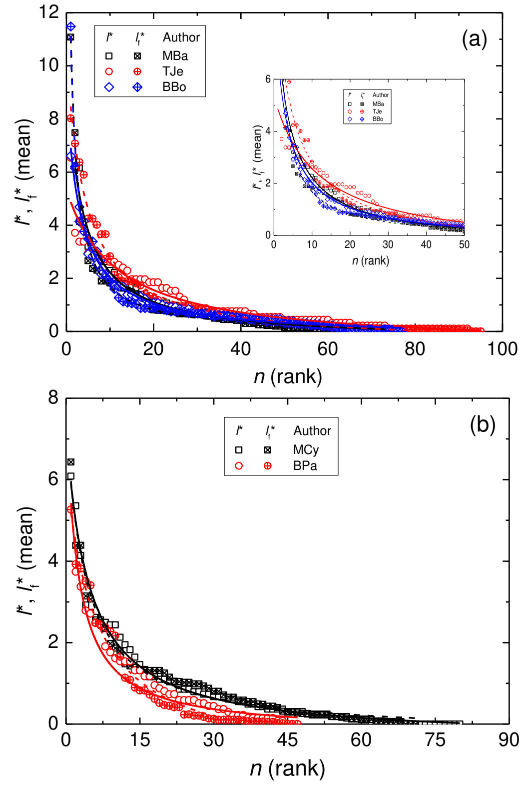

Figure 6: Plots of l^* and l_f^* of papers of different moderately-cited Polish authors against their rank n: (a) MBa, TJe BBo, and (b) MCy and BPa. Solid and dashed curves are drawn for l^* and l_f^* with the best-fit values of the parameters of Eq. (3) listed in Table 3. Inset in (a) shows magnified part of the plots close to the origin. Normalization of citations by mean citations.

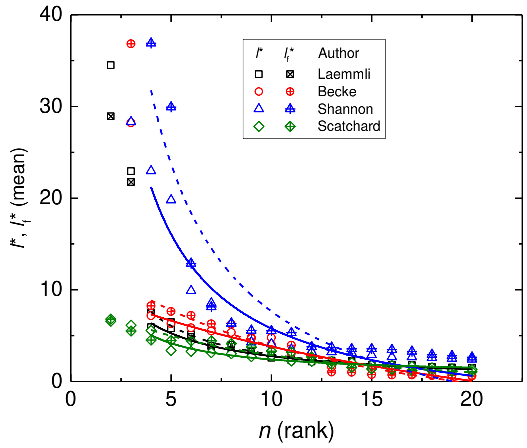

Figure 7: Plots of normalized citations l^* and l_f^* of papers of four highly-cited authors against their rank n in the range of $n < 20$. Solid and dashed curves are drawn for l^* and l_f^* in the range $4 \le n \le 20$ with the best-fit parameters of Eq. (3) listed in Table 4. Normalization of l^* and l_f^* citations by mean citations.

Figure 7 shows plots of relative citations l^* and l_f^* of papers of four highly-cited authors against their rank n in the range of $2 < n < 20$. Solid and dashed curves are

Table 3: Values of parameters of Langmuir-type function (3) for $l^*(n)$ and $l_f^*(n)$ data of different authors

Author	Citations l^* (with coauthors)				Citations l_f^* (without coauthors)			
	l_0^*	α	K	R^2	l_{f0}^*	α_f	K_f	R^2
Laemmli UK	23.40	1.0115	0.658	0.9841	111.7	1.0027	3.325	0.9946
Becke AD	44.81	1.0178	1.043	0.9292	1162	1.0007	27.82	0.9178
Southern EM	10.0	1.1083	0.105	0.8951	27.79	1.0174	0.624	0.9755
Felsenstein J	12.05	1.0730	0.161	0.9849	19.97	1.0404	0.30	0.9872
Shannon RD	523804	1.0	9997	0.9616	623514	1.0	9999	0.9679
Scatchard G	10.31	1.0617	0.244	0.9828	10.0	1.0874	0.20	0.9655
MBa	10.17	1.0362	0.307	0.9690	26.87	1.0041	1.152	0.9840
MCy	7.25	1.0550	0.214	0.9828	7.57	1.0342	0.258	0.9796
TJe	5.35	1.1083	0.088	0.9492	10.57	1.0505	0.233	0.9892
BBo	9.16	1.0255	0.315	0.9887	28.46	1.0014	1.551	0.9820
BPa	7.78	1.0294	0.415	0.9593	6.27	1.1571	0.171	0.9761
	10.0[a]	1.1857[a]	0.113[a]	0.9936[a]				

[a] Best-fit values for data without first point.

Table 4: Values of parameters of Langmuir-type function (3) for $l^*(n)$ data in the rank range $4 \le n \le 20$

Author	Citations l^* (with coauthors)				Citations l_f^* (without coauthors)			
	l_0^*	α	K	R^2	l_{f0}^*	α_f	K_f	R^2
Laemmli UK	33.57	1.0065	1.033	0.9591	177.8	1.002	5.383	0.9881
Becke AD	10.58	2.2264	0.040	0.9535	15.08	1.592	0.088	0.9438
Shannon RD	368560	1.0	3581	0.8991	640567	1.0	3987.4	0.8393
Scatchard G	8361	0.9999	472	0.9125	10	1.218	0.140	0.8760

Table 5: Values of parameters of Langmuir-type function (3) for $l_f^*(n_{eff})$ data

Author	N_c^{ef}	l_{f0}^{*ef}	α_f^{ef}	K_f^{ef}	R^2
Laemmli UK	32.83	165474	1.0	9993	0.9886
Becke AD	48.92	47087	1.00002	1248	0.8894
Southern EM	59.59	24632	1.00003	939.4	0.9556
Felsenstein J	75.31	26.05	1.03429	0.465	0.9853
Shannon RD	58.54	271559	1.0	7898	0.9144
Scatchard G	40.89	10.0	1.10088	0.340	0.9562
MBa	17.91	11.216	1.02460	1.727	0.9890
MCy	22.87	10.0	1.00177	1.341	0.8380
TJe	35.08	10.61	1.08863	0.369	0.9786
BBo	31.31	13.06	1.00222	1.718	0.9675
BPa	21.69	10.0	1.03760	0.747	0.7520

drawn for l^* and l_f^* in the range $4 \le n \le 20$ with the best-fit parameters of Eq. (3) listed in Table 4. From these plots the following features may be noted:

(1) The values of both l^* and l_f^* for all of the four authors show relatively poor fit in view of their irregular decreasing trends with increasing rank n.

(2) Except in the case of Scatchard, for low n the values of l^* and l_f^* steeply decreases with increasing n.

(3) In the case of Becke and Shannon, after an initial steep decrease the values of l^* and l_f^* decrease relatively poorly with increasing n.

These deviations of the data from the best-fit plots are associated with the difference in the number s_{max} of the active citation sites for the papers of an author (see assumption 1 in Section 2).

The distribution of relative citations l_f^* normalized by mean citations $<L_f>$ of N_c cited papers of different authors can also be analyzed as a function of their effective rank n_{eff}. Figure 8a and b shows the plots of relative citations l_f^* of papers of the 6 highly-cited and 5 moderately-cited authors, respectively, as a function of their effective rank n_{eff}. The curves in the figures are drawn according to Eq. (3) with the best-fit values of the new parameters l_{f0}^{*ef}, α_f^{ef} and K_f^{ef} of Eq. (3) listed in Table 5. Table 5 also includes the values of the effective number N_c^{ef} cited papers of different authors. It may be noted that the data for all authors are characterized by individual sets of the new parameters l_{f0}^{*ef}, α_f^{ef} and K_f^{ef} of Eq. (3) but the mutual dispersion among them is smaller when the values of l_f^* are normalized by the mean $<L_f>$ of cumulative fractions l_{nf} of citations of their papers. The values of the goodness-of-the fit parameter R^2 for the plots of the $l_f^*(n_{eff})$ data are also comparable with those for the plots of their $l_f^*(n)$ data (see Tables 3 and 4).

4.3. Parameters characterizing citation distributions of different authors

The above analysis of the publication output of different authors by Langmuir-type function (3) shows that real rank-order distributions of citations l_n and fractions l_{nf} of citations of papers of an author can be described by the two sets of three parameters: l_0, α and K, and l_{f0}, α_f and K_f, respectively. The parameters l_0 and l_{f0} are related to the parameters l_0^* and l_{f0}^* by the relations: $l_0 = l_0^*<L>$ and $l_{f0} = l_{f0}^*<L_f>$, where l_0^* and l_{f0}^* are given in Table 3

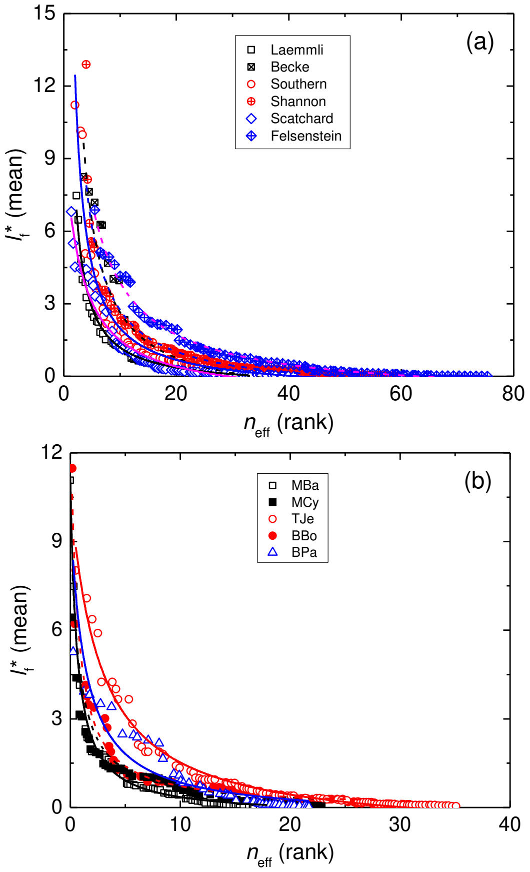

Figure 8: Plots of relatively citations l_f* of papers of (a) highly-cited and (b) moderately-cited authors against their effective rank n_{eff}. Curves are drawn with the best-fit values of the parameters of Eq. (3) listed in Table 4. Normalization of l_f* citations by mean citations.

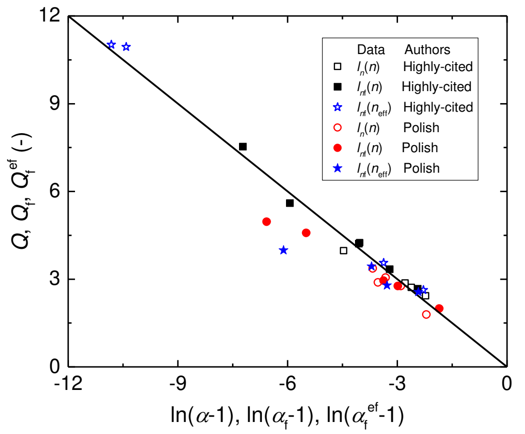

Figure 9: Plots of Q = ln$K$' = ln(KN_c) against ln(α–1), Q_f = lnK_f' = ln(K_fN_c) against ln(α_f –1) and Q_f^{ef} = lnK_f'^{ef} = ln($K_f^{ef}N_c^{ef}$) against ln(α_f^{ef}–1) for different authors. Linear plot represents the dependence expected from Eq. (10) with slope –1.

whereas <L> and <L_f> are listed in Table 2. Similarly, as in the case of the $l_{nf}(n)$ distributions, the $l_{nf}(n_{eff})$ distribution may be represented by a new set of three parameters: l_{f0}^{ef}, α_f^{ef} and K_f^{ef}, with l_{f0}^{ef} = l_{f0}*^{ef}<L_f>. The values of the new parameters l_{f0}*^{ef}, α_f^{ef} and K_f^{ef} are listed in Table 5 and those of <L_f> are given in Table 2.

The best-fit values of the parameters l_0, l_{f0} and l_{f0}^{ef} obtained from the $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$ distributions are usually very high and comparable (cf. Tables 2, 3 and 5). However, the mean citations <L> and <L_f> are parameters characteristic for an author. Similarly, the effectiveness parameters α, α_f and α_f^{ef}, the Langmuir constants K, K_f and K_f^{ef}, and the number N_c and the effective number N_c^{ef} of citable papers of an author are characteristic parameters of his/her publication output.

Eq. (7) may be rewritten in the form (cf. Eq. (5))

$$q = q_0 - \ln(\alpha - 1)\,, \tag{8}$$

where q = ln$k$, q_0 = lnk_0, k denotes dimensionless Langmuir constant K' = KN_c, K_f' = K_fN_c and K_f'* = K_f*N_c* as above, k_0 denotes the values of K', K_f' or K_f'* when (α–1) = 0, and α denotes the values of α, α_f and α_f* corresponding to K', K_f' and K_f'*, respectively. Now the parameter q represents the dimensionless differential energy Q, Q_f and Q_f* corresponding to KN_c, K_fN_c and K_f*N_c*, respectively. According to Eq. (8), q_0 = lnk_0 = 0 because the proportionality constant (kN_c)$_0$ = k_0 = 1 (see Section 3).

Figure 9 shows the data of Q = ln$K$' = ln(KN_c), Q_f = lnK_f' = ln(K_fN_c) and Q_f^{ef} = lnK_f'^{ef} = ln($K_f^{ef}N_c^{ef}$) against ln(α–1), ln(α_f–1) and ln(α_f^{ef}–1), respectively, for different authors according to Eq. (8). The values of different quantities were taken from Tables 1, 3 and 5. As expected from Eq. (8), one observes from Figure 9 a practically linearly decreasing plot of the data of Q, Q_f and Q_f*, with a slope of unity and an intercept of zero, against increasing ln(α–1), ln(α_f–1) and ln(α_f*–1), respectively, for the distribution of citations to the papers published by different authors. Since the effectiveness parameters α, α_f and α_f^{ef} are related to KN_c, K_fN_c and $K_f^{ef}N_c^{ef}$. by Eq. (8), it can be argued that the latter parameters, and the related dimensionless differential energies Q, Q_f and Q_f^{ef}, also characterize the publication output of an author.

From the above discussion it may be concluded that pairs of an author's mean citations <L> and dimensionless differential energy Q (where Q = ln(KN_c)) and <L_f> and Q_f (where Q_f = ln(K_fN_c)) are possible measures of characterization (and comparison) of his/her publication output with that of other authors from the distribution of cumulative citations l_n of their N_c cited papers published during their entire publication career. When contributions of coauthors to the citation l_n of the papers of an author are counted as l_{nf} and their ranking is arranged as effective rank n_{eff}, the relevant parameters obtained from the distribution of l_{nf} of their N_c^{ef} papers are: <L_f> and Q_f^{ef} = ln($K_f^{ef}N_c^{ef}$).

The main problem with the above analysis is that the pair of parameters cannot be combined into a single index for the comparison of the publication output of different authors.

4.4. Comparison of various single citation indices

Based on his stochastic model [55], Burrell [56] proposed a simple geometric distribution involving the number N of papers receiving L citations for an author to calculate his/her so-called quasi h index (h'). Using the published data of the number of papers N and citations L of 15 authors, Burrell [56] showed that, if one or more outliers in the upper tail of the citation distribution of an author are omitted, his approach gives good estimate of their h index. It may be noted that the values of this h' index, estimated by Burrell, of the 15 authors are comparable with the values of their citation radius $R = (L/\pi)^{1/2}$ (cf. ref. [57]).

From the distributions of citations $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$ data of the different authors their Hirsch index h, the new index h_f and the Hirsch-type index h_m, respectively, were determined using the procedures mentioned above. Following the proposal of Batista et al. [18], the so-called h_1 index was calculated from the relation: $h_1 = h^2/A_h$, where A_h is the total number of coauthors of a particular author in the h-core. Excluding the citations of the outliers, the citation radii R and R_f of the authors were also calculated from their cumulative citations L and L_f (cf. ref. [57]). The calculated values of the h, h_f, h_m and h_1 indices and the citation radii R and R_f of the authors are included in Table 1. The values of the number A_h of authors in the h-core, the number of cumulative citations L of the authors without their outliers, the number N_c of papers without outliers, and the number N_m of the papers involved in the h_m index of the authors are given in Table 1 in the parentheses.

From Table 1 it may be noted that the value of the h index of different authors is related directly neither to their cumulative citations L nor to the number of papers published by them. For example, Felsenstein, with L = 46478 citations, has an h index equal to 51 whereas Becke, with citations twice the citations of Felsenstein, has h equal to 48 only. Another example is that of TJe and MCy having h index equal to 18 but the latter's papers received citations about 40% higher than that of the former. Similarly, Becke with 70 cited papers has a higher h index than that of Southern and Shannon with 110 and 152 cited papers, respectively.

As in the case of the h index of different highly-cited authors, their h_f index is also related neither to their cumulative citations L_f nor to the number of papers published by them (Table 1). For example, Laemmli and Becke have $h_f = 40$ but Becke has L_f twice higher than that of Laemmli. Similarly, although Shannon has about 40% higher cited papers than Southern, both have the same $h_f = 28$. In the case of moderately-cited Polish professors the above trends are not so clear. However, as judged from the number A_h of coauthors in the h-core and the number N_c of cited papers of an author, the value of his/her h_f index is determined by the number of coauthors of the high-ranked (i.e. more cited) papers and the number of his/her published papers.

As seen from Table 1, the estimated citation radius R for different moderately-cited Polish authors are comparable with their h index. Similarly, the values of R_f for these authors are also comparable with their h_f index. In contrast to this, the estimated values of R for the highly-cited authors are much higher than their h index even when citations from outliers are excluded from cumulative citations. However, the values of their R_f are close to their h_f in some cases or somewhat higher than their h_f in other cases. These observations suggest that $R \approx h$ and $R_f \approx h_f$ when the distribution of citations l_n and l_{nf} of succeeding papers in the range of more cited papers (i.e. $n < h$) is a smoothly, rather than a steeply, decreasing function of lowering rank n of the papers. This is the situation in the case of moderately-cited Polish professors. However, the values of R and R_f are higher than h and h_f, respectively, when the papers of rank $n < h$ in the citation distributions receive relatively high citations l_n and l_{nf}. This is the situation in the case of highly-cited authors (see Table 1). While calculating his h' index of different authors from their cumulative citations L and papers N, Burrell [56] also attributed the estimated high values of the h' index of some of the authors than those of their h index to the citations of highly cited papers.

Table 1 reveals that the h_1 index is not a suitable citation-based measure of comparison of the research output of different authors. For example, Laemmli, with citations five times higher than that of Felsenstein, has h_1 roughly one-half of the h_1 of Felsenstein. Similarly, although MBa's papers received more citations than MCy's papers, MBa has h_1 lower than that of MCy by a factor of 7.5. This anomalous behavior of the h_1 index is associated with the total number A_h of coauthors considered in calculating it from the h index. A similar behavior was reported by Schreiber [42].

It is interesting to note the difference in the number of papers included in the calculation of citation radii R and R_f and Hirsch and Hirsch-type indices h, h_f and h_m of different authors. While the citations received by all cited papers N_c of an author are taken into account in the calculation of his/her citation radii R and R_f, counting of h index is based on cited papers $n_c \le h < N_c$ and that of h_f index is based on even a lower number of cited papers (see Table 1). For example, in the case of Laemmli, the counting of his h index involves 58 papers but that of h_f index based on fractions of citations of his papers with coauthors is just 40. The deficiencies of the h and h_f indices of different authors mentioned above are overcome by the h_m index, which is determined by fractions n_{eff} of the rank of a higher number of their papers. The determination of this h_m index of an author involves the number of papers comparable with that involved in determining his/her h index (see Table 1). In the case of Laemmli for example, the number N_m of the cited papers involved in the calculation of h_m index is 55 in comparison with $n_c = 58$ cited papers in determining the h index. These observations suggest that h_m index is superior to h and h_f for comparison of the publication output of different authors because it takes into account the number of coauthors of their papers as well as the number of cited papers comparable with those involved in the calculation of the classical h index.

From the data of cumulative fraction L_f of citations of the authors given in Table 1 one finds empirically that, with the exception of Felsenstein and MBa, $h_m \approx (2L_f)^{1/3}$. The estimated value of h_m of 26.70 is much lower than the calculated 41.92 for Felsenstein whereas

Table 6: Comparison of different scaling parameters for citations analysis of various authors

Author	t (yr)	N_c	<L>	<L_f>	b	b_f	b_m
Laemmli UK	45	84	166.40 (2)	65.99 (3)	1.29 (2)	0.89 (2)	0.476 (4)
Becke AD	36	70	232.88 (1)	178.41 (1)	1.33 (1)	1.11 (1)	0.8728 (2)
Southern EM	56	110	60.66 (5)	31.75 (4)	0.73 (7)	0.50 (6)	0.341 (7)
Felsenstein J	48	99	130.98 (3)	101.21 (2)	1.06 (5)	0.625 (4)	0.8733 (1)
Shannon RD	45	152	41.66 (6)	15.98 (6)	1.00 (6)	0.622 (5)	0.407 (5)
Scatchard G	68	73	78.54 (4)	38.90 (5)	0.485 (10)	0.368 (8)	0.236 (9)
MBa	15	70	12.19 (10)	3.39 (11)	1.133 (4)	0.467 (7)	0.349 (6)
MCy	35	68	16.43 (7)	4.56 (9)	0.514 (8)	0.257 (10)	0.230 (10)
TJe	15	92	9.17 (11)	4.24 (10)	1.20 (3)	0.667 (3)	0.537 (3)
BBo	41	74	13.01 (9)	5.32 (8)	0.415 (11)	0.244 (11)	0.203 (11)
BPa	32	44	13.79 (8)	9.69 (7)	0.50 (9)	0.344 (9)	0.3225 (8)

that of h_m of 7.80 is higher than the calculated 5.24 for MBa. As seen from Table 1, these anomalies are associated with differences in the number of coauthors of the top-cited papers. The number of coauthors per paper in the former case is relatively low in comparison with those in the latter case.

Finally, it should be noted from Tables 1 and 2 that the h, h_f and h_m indices of different authors are, in general, related to their mean citations <L> and <L_f>. Consequently, all highly-cited authors with high <L> and <L_f> are in the top six positions whereas the moderately-cited authors with the low of <L> and <L_f> occupy positions lower than those of the highly-cited authors.

4.5. Scaling parameters for research output of different authors

The h, h_f and h_m indices and the pairs of characteristic parameters <L> and Q, and <L_f> and Q_f of an author, discussed above, are related to the distributions $l_n(n)$, $l_{nf}(n)$ and $l_{nf}(n_{eff})$, respectively, constructed from his/her bibliometric data. The inherent disadvantage with all of these measures of the publication output of an author is that their values do not take into account the length of publication career of the author. It is difficult to calculate career-length independent characteristic parameters from Langmuir-type function for individual authors, but this factor can be considered in the h, h_f and h_m indices using the ideas used before [58].

The relationship between cumulative citations $L(t)$ and Hirsch index h is given by [1]

$$L(t) = Ah^2(t), \tag{9}$$

whereas according to progressive nucleation mechanism [57,58] and stochastic model [55]

$$L(t) = at^2. \tag{10}$$

In the above equations A is an empirical constant lying between 3 and 5, and a is the so-called citation acceleration related, among others, to the average number of papers published per year by an author. From the above relations one obtains

$$h = bt, \tag{11}$$

where the proportionality constant $b = (a/A)^{1/2}$, with units year^{-1}.

Relation (11) means that the average h of an author increases linearly with publication career t. Since other Hirsch-related indices like h_f and h_m are based on similar concept, it can be argued that these indices also increase linearly with t. The calculated values of the parameter b, denoted as b, b_f and b_m corresponding to h, h_f and h_m, respectively, are given in Table 6.

From Table 6 it may be noted that the parameters b, b_f and b_m for different authors lead to a major change in their ranking. When the contribution of coauthors is taken into account for a comparison of the publication of different authors, the ranking of highly-cited authors is drastically changed, and highly-cited authors like Southern EM and Scatchard G are shifted downward but moderately-cited authors like TJe and MBa of short publication career are shifted upward to 3rd and 6th ranks, respectively.

5. Summary and conclusions

The real data of cumulative citations l_n of nth paper of individual N papers published by selected authors are analyzed without and with consideration of the number A_n of coauthors of the paper, the normalization of citations l_n and cumulative fraction l_{nf} of citation of the nth paper by mean and median citations of the citations l_n of all N_c cited papers and the determination of cumulative fraction n_{eff} of the rank of the l_{nf} citations. In view of different authorship patterns used in the lists of authors of the papers considered in this study and enormously high number of coauthors in many cases, l_{nf} and n_{eff} were calculated on the assumption of equal contributions of A_n coauthors of the nth-ranked paper fetching l_n citations.

It was found that Langmuir-type function (3) is not adequate for comparison of the publication output of different authors because it describes the rank-order distribution patterns satisfactorily in terms of two parameters. However, when equal contribution of coauthors is taken into account for the comparison of their publication output, the h_m index is more consistent than other Hirsch-type indices. To account for the length of their publication career t, it is suggested to use scaling parameters h/t, h_f/t and h_m/t for a comparison of the publication output of different authors

Acknowledgement

The author expresses his gratitude to Dr Kazimierz Wójcik for his assistance with collection of the bibliometric data used in this work.

References

[1] J.E. Hirsch, An index to quantify an individual's scientific research output, Proceedings of the National academy of Sciences of the United States of America 102(46) (2005) 16569-16572.

[2] M. Kosmulski, A new Hirsch-type index saves time and works equally well as the original h-index, ISSI Newsletter 2(3) (2006) 4-6.

[3] B. Jin, L. Liang, R. Rousseau, L. Egghe, The R- and AR-indices: complementing the h-index, Chinese Science Bulletin 52(6) (2007) 855-863.

[4] S. Alonso, F. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-index: A review focused in its variants, computation and standardization for different scientific fields, Journal of Informetrics 3(4) (2009) 273-280.

[5] T.R. Anderson, R.K.S. Hankin, P.D. Killworth, Beyond the Durfee square: Enhancing the h-index to score total publication output, Scientometrics 76(3) (2008) 577-588.

[6] Q.L. Burrell, On Hirsch's h, Egghe's g and Kosmulski's h(2). Scientometrics 79(1) (2009) 79-91.

[7] L. Egghe, Characteristic scores and scales based on h-type indices. Journal of Informetrics 4(1) (2010) 14–22.

[8] F. Franceschini, D. Maisano, The Hirsch spectrum: a novel tool for analyzing scientific journals, Journal of Informetrics 4(1) (2010) 64-73.

[9] F. Franceschini, D. Maisano, The citation triad: an overview of a scientist's publication output based on Ferrers diagrams, Journal of Informetrics 4(4) (2010) 503-511.

[10] W. Glänzel, A. Schubert, Hirsch-type characteristics of the tail of distributions. The generalized h-index, Journal of Informetrics 4(1) (2010) 118-123.

[11] A. Tietze, P. Hofmann, The h-index and multi-author h_m-index for individual researchers in condensed matter physics, Scientometrics 119 (2019) 171-185.

[12] M. Ameer, M.T. Afzal, Evaluation of h-index and its qualitative and quantitative variants in Neuroscience, Scientometrics 121 (2019) 653-673.

[13] S. Li, H. Shen, P. Bao, X, Cheng, h_u-index: a unified index to quantify individuals across disciplines, Scientometrics 126 (2021) 3209-3226.

[14] M. Kosmulski, New seniority-independent Hirsh-type index, Journal of Informetrics 3(4) (2009) 341-347.

[15] H.A. Abt, A publication index that is independent of age, Scientometrics 91(3) (2011) 1053-1058.

[16] R. Mannella, P. Rossi, On the time dependence of the h-index, Journal of Informetrics 7(1) (2013) 176-182.

[17] D. Pradhan, P.S. Paul, U. Maheswari, S. Nandi, T. Chakraborty, C3-index: a PageRank based multi-faceted metric for authors' performance measurement, Scientometrics 110 (2017) 253–273.

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

[19] M. Ausloos, A simple law about co-authors and their ranking: The co-author core, Scientometrics 95(3) (2013) 895-909

[20] D.D.S. Price, Multiple authorship, Science 212(4498) (1981) 986.

[21] S.E. Hodge, D.A. Greenberg, Publication credit, Science 212(4498) (1981) 950.

[22] P. Vinkler, Research contribution, authorship and team cooperativeness, Scientometrics 26(1) (1993) 213-230.

[23] A.F. Pereira de Araújo, Increasing discrepancy between absolute and effective indexes of research output in a Brazilian academic department, Scientometrics 74(3) (2008) 425-437.

[24] N. Assimakis, M. Adam, A new author's productivity index: p-index, Scientometrics 85(2) (2010) 415-427.

[25] R.S.J. Tol, Credit where credit's due: accounting for co-authorship in citation counts, Scientometrics 89(1) (2011) 291-299.

[26] S. Sahoo, Analyzing research performance: proposition of a new complementary index, Scientometrics 108(2) (2016) 489-504.

[27] J.-W. Hsu, D.-W. Huang, Correlation between impact and collaboration, Scientometrics 86(1) (2011) 317-324.

[28] M. Kosmulski, The order in the lists of authors in multi-author papers revisited, Journal of Informetrics 6(4) (2012) 639-644.

[29] C.H. Sekercioglu, Quantifying coauthor contributions, Science 322(5900) (2008) 371.

[30] N.T. Hagen, Credit for coauthors, Science 323(5914) (2009) 583.

[31] N.T. Hagen, Harmonic publication and citation counting: sharing authorship credit equitably – not equally, geometrically or arithmetically, Scientometrics 84(3) (2010) 785-793.

[32] N.T. Hagen, Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy, Journal of Informetrics 7(3) (2013) 784-793.

[33] C.T. Zhang, A proposal for calculating weighted citations based on author rank, EMBO Reports 10(5) (2009) 416-417.

[34] I. Podlubny, Comparison of scientific impact expressed by the number of citations in different fields of science, Scientometrics 64(1) (2005) 95-99.

[35] J.E. Iglesias, C. Pecharroman, Scaling the h-index for different scientific ISI fields, Scientometrics 73(3) (2007) 303-320.

[36] J. Lundberg, Lifting the crown – Citation z-score, Journal of Informetrics 1(2) (2007) 145-154.

[37] F. Radicchi, S., Fortunado, C. Castellano, Universality of citation distributions: Towards an objective measure of scientific impact, Proceedings of the National academy of Sciences of the United States of America 105(45) (2008) 17268–17272.

[38] G. Abramo, T. Cicero, C.A. D'Angelo, Revisiting the scaling of citations for research assessment, Journal of Informetrics 6(3) (2012) 420-479.

[39] L. Bornmann, H.-D. Daniel, Universality of citation distribution – a validation of Radicchi et al.'s relative indicator $c_f = c/c_0$ at the micro level using data from chemistry, Journal of the American Society for Information Science and Technology 60(8) (2009) 1664-1670.

[40] J.E. Hirsch, An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship, Scientometrics 85(3) (2010) 741-754.

[41] L. Egghe, Mathematical theory of the h- and g-index in case of fractional counting of authorship, Journal of the American Society for Information Science and Technology 59(10) (2008) 1608-1616.

[42] M. Schreiber, A modification of the h-index: The h_m-index accounts for multi-authored manuscripts, Journal of Informetrics 2(3) (2008) 211-216.

[43] M. Schreiber, A case study of the modified Hirsch index h_m accounting for multiple co-authors, Journal of the

American Society for Information Science and Technology 60(6) (2009) 1274-1282.
[44] G. Abramo, C.A. D'Angelo, F. Rosati, The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences, Journal of Informetrics 7 (2013) 198-208.
[45] M. Dunaiski, J. Geldenhuys, W. Visser, Author ranking evaluation at scale, Journal of Informetrics 12 (2018) 679-702.
[46] M. Salman, M.M. Ahmed, M.T. Afzal, Assessment of author ranking indices based on multi-authorship, Scientometrics 126 (2021) 4153-4172.
[47] L. Egghe, A rationale for the Hirsch-index rank-order distrubtion and a comparison with the impact factor rank-order distribution, Journal of the American Society for Information Science and Technology 60(10) (2009) 2142-2144.
[48] L. Egghe, R. Rousseau, An informetric model for the Hirsch-index, Scientometrics 69(1), (2006) 121-129.
[49] L. Egghe, R. Rousseau, The Hirsch index of a shifted Lotka function and its relation with the impact factor, Journal of the American Society for Information Science and Technology 63(5) (2012) 1048-1053.
[50] K. Sangwal, Comparison of different mathematical functions for the analysis of citation distribution of papers of individual authors, Journal of Informetrics 7(1) (2013) 36-49.
[51] Q.L. Burrell, Formulae for the h-index: A lack of robustness in Lotkaian informetrics?, Journal of the American Society for Information Science and Technology 64(7) (2013) 1507-1514.
[52] K. Sangwal, Citation and impact factor distributions of scientific journals published in individual countries, Journal of Informetrics 7(3) (2013) 487-504.
[53] K. Sangwal, Distributions of citation of papers of individual authors publishing in different scientific disciplines: Application of Langmuir-type function, Journal of Informetrics 8(4) (2014) 972-984.
[54] K.Y. Chuang, Y.S. Ho, Bibliometric profile of top-cited single-author articles in the Science Citation Index Expanded, Journal of Informetrics 8(4) (2014) 951-962.
[55] Q.L. Burrell, Hirsch's h-index: A stochastic model, Journal of Informetrics 1(1) (2007) 16-25.
[56] Q.L. Burrell, The h-index: A case of the tail wagging the dog?, Journal of Informetrics 7(3) (2013) 774-783.
[57] K. Sangwal, On the relationship between citations of publication output and Hirsch index *h* of authors: conceptualization of tapered Hirsch index h_T, circular citation area radius *R* and citation acceleration *a*, Scientometrics 93(3) (2012) 987-1004.
[58] K. Sangwal, On the age-independent publication index, Scientometrics 91(3) (2012) 1053-1058.