Hirsch Index Rankings Require Scaling and Higher Moment

Aggelos Bletsas[1*] and John N. Sahalos[1]

[1] Radio-Communications Laboratory (RCL), Department of Physics

Aristotle University of Thessaloniki, Thessaloniki, 54453 Greece

tel. +30-2310-998161, fax. +30-2310-998069

[*]corresponding author: aggelos@media.mit.edu, bletsas@auth.gr

Abstract: Simple bibliometric indicators, such as average number of citations per publication per researcher, or the recently-proposed Hirsch index (h-index), are nowadays tracked by online repositories including Web-of-Science (WOS) and often affect critical decision making. This work proposes appropriate scaling of the h-index, based on its probability distribution which is calculated for *any* underlying citation distribution. The proposed approach outperforms existing index estimation models which have focused on the expected value only (i.e. first moment). Furthermore, it is shown that average number of citations per publication per scientific field, total number of publications per researcher, as well as researcher's h-index measured value, expected value and standard deviation constitute the minimum information required for meaningful h-index ranking campaigns; otherwise contradicting ranking results emerge. This work may potentially shed light to (current or future) large-scale, h-index-based bibliometric evaluations.

Keywords: Hirsch index, citation networks, scientific activity/visibility, bibliometrics.

1. Introduction

Simple bibliometric indices, such as total number of publications or average citations per publication, are commonly exploited in evaluation of individual scientists (Bornmann, Mutz and Daniel, 2008) (e.g. recruitment, grant approval), research groups (van Raan, 2006) or whole academic and research institutions (e.g. annual university rankings worldwide) (THES, 2008). Given the plethora of publication venues the last decade, mainly due to online, open access journals, assessment of bibliometric information for academic/research decision making becomes challenging.

This work spearheads ranking research according to simple bibliometric information such as total number of journal publications N_T and citations C_T per researcher respectively, average number of citations per publication per researcher C_T / N_T, and the recently proposed Hirsch index (h-index for brevity) that depends on number of publications (activity) as well as number of citations (visibility) (Hirsch 2005, 2007); all such information is readily available by commercial databases, such as Thompson Scientific Web-of-Science (WOS) and Elsevier Scopus or can be calculated through other online repositories.[1]

The contribution of this work is summarized below:

1. The probability distribution of the h-index is calculated for *any* underlying citation distribution. Thus, all h-index moments could be in principle calculated. This is in sharp contrast to prior art which has focused on the first moment only (expected value).

2. Experimental, easy-to-reproduce study is provided, demonstrating that a simple model of Tsallis citation distribution provides more accurate estimates of the h-index, compared to Weibull (also known as Stretched Exponential) or Power Law/Pareto distributions. The

---

[1] e.g. the "Publish or Perish" software tool (http://www.harzing.com/pop.htm) provides bibliometric indicators including the h-index, using Google Scholar. It is noted that assesing citation indexing thoroughness/competence of online repositories is well beyond the scope of this work.

latter have been heavily used in bibliometric studies. This finding confirms recent reports on the preference of Tsallis distribution (also known as q-exponential or type-II generalized Pareto) as a better model for the citation distribution.

3. Knowledge of the h-index probability distribution provides simple computation of *both* h-index expected value *and* standard deviation. It is shown that average number of citations per publication per scientific field, total number of publications per researcher, as well as researcher's h-index measured value, expected value and standard deviation constitute the minimum information required for meaningful h-index ranking campaigns; otherwise contradicting ranking results emerge.

Section 2 provides calculation of the h-index probability density function. Section 3 offers experimental validation of the proposed h-index model, utilizing easy-to-reproduce WoS measurements. Section 4 provides the ranking case study and finally, Section 5 provides conclusion and future work directions.

2. H-index modelling

The h-index denotes the maximum number of publications among a {scientist, journal or institution} collection of publications with citations each, at least equal to this number (e.g. h-index of 9, in a collection of 20 publications, indicates that there are 9 publications in the 20-publication set, with *at least* 9 citations for each publication, while all the rest 11 publications have *at most* 9 citations each). Assuming *continuous* random variable X for the number of citations per publication per scientific field (citation distribution), the h-index can be directly derived from the complementary cumulative distribution function $F_c(x) = \Pr\{X > x\}$ of X, with two different methods.

2.1 Expected value of h-index or single-moment method (Method 1)

The first method provides an estimate of the average value of h-index (first moment); if Y_κ denotes the number of citations of the κ_{th} most cited publication ($Y_1 \geq Y_2 \geq ... \geq Y_\kappa \geq ... \geq Y_{N_T}$) among a set of N_T publications, then the following holds:

$$F_c(Y_\kappa) = \Pr\{X > Y_\kappa\} \cong \frac{\kappa}{N_T}, \text{ ``large'' } N_T \quad (1)$$

The second equality stems from order statistics and holds for ``sufficiently large'' N_T, according to the law of large numbers (i.e. the ratio κ/N_T equals probability).

The expression above has appeared in the literature, in the context of Zipf laws and relevant studies regarding the κ_{th} most cited researcher (Laherrere and Sornette (1998), page 528, immediately below Eq. (5), where $P_c = \kappa/N_T \equiv F_c(Y_\kappa)$ is provided). Solving for Y_κ in Eq. (1) and equating $Y_h = h$ provides the expected value of the h-index, given that the latter indicates the highest number of publications with citations at least equal to this number. An expression similar to the above, using Gumbel's characteristic extreme values, has been tested for Power Law/Pareto citation distributions i.e. $F_c(x) \propto x^{-(\alpha-1)}$ (Schubert and Glänzel (2007)). It is noted that the Power Law distribution has been extensively utilized in bibliometrics (e.g. see Redner (1998) and references therein or Egghe and Rousseau (2006), for h-index modelling with Lotkaian systems governed by a power law).

Eq. (1) with $Y_h = h$ shows that the measured ratio h/N_T of a researcher provides an estimate of the probability her/his publication receives at least h citations. Therefore, only scientists with equal measured h-index, or equal N_T are directly comparable, otherwise appropriate scaling is required.

2.2 Probability distribution of h-index or all-moments method (Method 2)

The proposed second method offers the probability distribution of the h-index and as such, the standard deviation as well as higher moments, apart from the average value. This way, one could assess how much (i.e. how many standard deviations σ_h) a measured h-index value is above or below the expected value, for given total number of publications N_T and average number of citations per publication per scientific field $E\{X\}$.[2] In principle, h-index is an integer within $\{0, 1,\ldots, N_T\}$ and equals n, when n publications out of N_T have received at least n citations each, and there are no $n+1$ publications with at least $n+1$ citations each; otherwise, the h-index would be $n+1$ and not n. The above is written probabilistically as follows:

$$\Pr\{h = n, \varepsilon\} = \binom{N_T}{n}[F_c(n)]^n[1 - F_c(n+\varepsilon)]^{(N_T-n)}, \ n = \{1, 2,\ldots, N_T\}, \ \forall \varepsilon \in [0,1)$$

$$\Pr\{h = 0\} = [1 - F_c(1)]^{N_T}, \tag{2}$$

where $\binom{N_T}{n}$ denotes the combinations of n out of N_T and the second line corresponds to the probability that all N_T publications receive no citation.

It is noted that the first term $[F_c(n)]^n$ denotes the probability that n publications receiving *at least n* citations each, while the second term $[1 - F_c(n+\varepsilon)]^{(N_T-n)}$ denotes the probability that $(N_T - n)$ publications receiving *at most* $(n + \varepsilon)$ citations each, where ε is a non-negative constant, strictly less than unity. For any such parameter $\varepsilon \in [0,1)$, the definition of h-index is satisfied (``can't find $h+1$ publications with at least $h+1$ citations each''). For the same reason, the first term $[F_c(n)]^n$ should not be equal to $[F_c(n+\varepsilon)]^n$, since in the latter case, the minimum number of citations in the h-defining core would be greater than h, violating the definition (e.g. a group

---

[2] The operator E{x} denotes the expected value of x.

of 20 publications with 9 citations each would have h-index strictly greater than 9 and not exactly 9, as it should according to the definition). Parameter ε can be calculated according to the total probability law $\Pr\{h=0\} + \sum_{n=1}^{N_T} \Pr\{h=n,\varepsilon\} = 1$.

With this method, all moments $\mathrm{E}\{h^m\}$ of the h-index can be computed ($\mathrm{E}\{h^m\} = \sum_{n=1}^{N_T} n^m \Pr\{h=n,\varepsilon\}$), providing average value (m=1), as well as variance $\sigma_h^2 = \mathrm{E}\{h^2\} - \left(\mathrm{E}\{h\}^2\right)$ (m=2). Knowledge of standard deviation σ_h allows for appropriate scaling, at the cost of computation requirements for possibly large-N_T factorials in $\begin{pmatrix} N_T \\ n \end{pmatrix}$.

It is further noted that method 2 assumes that the number of citations per publication X (for a given scientific field) is a *continuous* random variable, while h-index is modeled as a non-negative integer (*discrete* random variable). For this reason, as well as the fact that parameter $\varepsilon \in [0,1)$ is utilized, method 2 could introduce a quantization error, on the order of one half (0.5). Given that h-index spans values from 0 to N_T (which is usually a two-digit number for the average researcher), quantization error on the order of 0.5 becomes trivial, for all possible values of h-index. In fact, Section 2.3 shows that numerical results difference between methods 1 and 2 is negligible. Ideas of extending the h-index to the continuous (real) case already exist in the recent literature; e.g. see work by Ruane and Tol (2008), with rational h-index for group of researchers, Schreiber (2008) where interpolation is used, as well as work by Chai, Hua, Rousseau and Wan (2008), Guns and Rousseau (2009), where rational variants are also provided. In this work, we keep the original definition (according to which h-index is integer) and calculate its probability distribution in Eq. (2).

2.3 Discussion

Table 1 displays three examples of empirical, two-parameter citation distributions. Ideas on how such two parameters can be practically set given measurements, are provided in Section 4.1. Table 1 shows that the two parameters define the expected number of citations per publication $E\{X\}$ in a given scientific field; thus, knowledge of $E\{X\}$, as well as knowledge of one of the two parameters clearly defines the citation distribution. Specifically, Table 1 provides continuous probability density $f(x)$ for the citation distribution of random variable X (citations per publication for a given scientific field), respective complementary cumulative distribution function $F_c(x)$, expected value $E\{X\}$ and ordered Y_κ's (defined in Section 2.1).[3] The chosen distributions have been utilized in bibliometrics to model empirical data (Redner, 1998), (Laherrere and Sornette, 1998), (Tsallis and de Albuquerque, 2000). It is further noted that Tsallis distribution is also known as q-exponential distribution or generalized type II Pareto distribution (Shalizi, 2007 and references therein), (Burrell, 2008). In order to avoid confusion with the standard Power Law/Pareto distributions, we keep the term ``Tsallis distribution''.

As a first example, we calculate the expected value for the h-index, using the first citation distribution of Table 1 (Power Law):

$$E\{h\} = \left[\frac{a-2}{a-1}\right]^{\frac{a-1}{a}} \left[E\{X\}\right]^{\frac{a-1}{a}} N_T^{1/a}, \quad a > 2. \qquad (3)$$

For the more general case of Pareto distributions ($F_c(x) \propto x^{-(\alpha-1)}$), among which Power Law is a special case), the h-index becomes:

$$E\{h\} = b\left[E\{X\}\right]^{\frac{a-1}{a}} N_T^{1/a}, \quad a > 2, \qquad (4)$$

where b is a constant. Eq. (4) with parameter $a' = a-1$ was reported by Schubert and Glänzel

---

[3] It is remarked that $\Gamma(x)$ in Table 1 denotes the natural logarithm and Gamma function, respectively.

(2007). Paretian distributions are convenient for method 1, given that they can provide closed-form solution, shown above. This is not the case for other empirical citation distributions, such as Tsallis and Weibull, where equation $Y_h = h$ must be solved numerically and no closed-form expression exists.

Figs. 1, 2 depict the calculated average h-index as a function of average citations per publication $E\{X\}$ (for a given scientific field) and total number of publications N_T, for Tsallis with parameter $q = \{1.5, 1.6\}$ and Weibull with parameter c = $\{0.45, 0.35\}$ respectively. Tsallis with $q = 1.53$ has modelled the citation distribution of publications, appeared in 1981 and cited between 1981-June 1997, while $q = 1.64$ has modelled the citation distribution of papers published between 1975-1994 and cited as of June 1997 (Tsallis and de Albuquerque, 2000). The same data sets have been modelled by Weibull with $c = 0.44$ and $c = 0.39$ respectively, implying that two-parameter functions could describe citation distributions of different measurement (time) periods. Similar modelling with Weibull can be found in (Laherrere and Sornette, 1998). It is also remarked that Weibull (also known as Stretched Exponential) was instead used in (Hirsch, 2005) and (Iglesias and Pecharroman, 2007) to model the ordered Y_κ's of Eq. (1) and not the citation distribution itself.

Figs. 1, 2 show that the two methods for h-index estimation provide similar results, as expected; the numerical results difference on the order of 0.2 (or less) is barely noticeable and is attributed to the inherent quantization noise of method 2 (as explained before). It is clear that h-index heavily depends on number of papers N_T as well as field of research; the latter affects average number of citations per publication $E\{X\}$. See for example, the case of $E\{X\} = 10$ citations per publication, for Tsallis and $q = 1.6$ in Fig. 1: increasing the number of publications from 20 το 60 approximately doubles the h-index. This may explain the phenomenon that h-index increases almost linearly with time, as recently observed by Hirsch (2007) and

theoretically justified by Burrell (2007) and Egghe (2007). It also underlines the fact that h-index strongly depends on field of research (E{X}), as well as total number of publications (N_T), i.e. scientists (in the same scientific field) with different number of publications are not directly comparable in terms of the h-index, unless such difference in N_T is explicitly taken into account (Section 4).

3. H-index model validation

Both methods of Eqs. (1), (2) suggest that h-index heavily depends on the scientific field and its underlying citation distribution, as well as total number of papers. In order to assess both the qualitative and quantitative validity of the above h-index calculation models, a simple experiment was conducted. Web of Science (WoS) was utilized to measure total number of publications N_T and their respective h-index h_{meas}, total citations C_T or citations excluding self-citations C_0, gathered within well-defined time intervals for well-known Institute of Electrical and Electronics Engineers (IEEE) Transactions, relevant to wireless communications (Table 2). These journals are considered within the top publication venues in their respective scientific domains.

It was a striking surprise to see how closely the measured h-index from WoS (h_{meas}) matches the estimated h-index (h_{est}), with Tsallis citation distribution, while the two others (Power Law/Pareto and Weibull) provided increased estimation errors. See for example, the first raw of Table 2: there were 182 publications in 1993 at the IEEE Transactions of Information Theory. Until December 2008, those publications received 4505 citations, 2941 citations excluding self-citations and achieved an h-index of 32, i.e. 32 papers out of 182 received at least 32 citations, while all the rest received at most 32 citations. The average number of citations per

paper for the field of information theory is estimated to $E\{X\} = C_T/N_T = 4505/182 = 24.75$. For that value of $E\{X\}$ and $N_T = 180$ publications, Fig. 1-right estimates h-index to $E\{h\} \approx 31$, assuming Tsallis Distribution with $q = 1.6$; the estimated average h-value is very close to the measured value of 32. Similarly, Fig. 2-right provides an estimate close to 28.5, assuming Weibull Distribution with $c = 0.35$. Moreover, Eq. (3) provides h-index equal to 30.32 assuming Power Law citation distribution with $\alpha = 3$ (and $N_T = 182$, $E\{X\} = 24.75$).

Addressing a more recent dataset, it is observed that 307 papers were published in 2003 IEEE Transactions on Communications and achieved an h-index of 30, by December 2008. The calculated h-index value for the reported value of $E\{X\} = 18.25$ is approximately 33.5, according to Fig. 1-right and 29.46 according to Power Law of Eq. (3), with $\alpha = 3$. This finding suggests that citation distribution parameters depend on the citation window duration, i.e. parameters appropriate for a 16-year citation window are slightly different than those for a 6-year window.

Last line of Table 2 shows that $N_T = 5736$ journals, published within 1993-2007 in IEEE Transactions on Signal Processing, produced $C_T = 87070$ citations and measured h-index equal to 98, as of Dec. 2008. For $N_T = 5736$ and $E\{X\} = 87070/5736 = 15.18$, the calculated h-index for Tsallis citation distribution with $q = 1.6$, $q = 1.5$ becomes 102.86, 99.87 respectively, very close to the measured value of 98 provided by WoS. On the contrary, the calculated h-index for Power Law of Eq. 3 with $\alpha = 3$ becomes 69.13, while for Weibull citation distribution with $c = 0.35$, $c = 0.45$ becomes 133.12, 122.33 respectively, all considerably different than the measured value of 98.

Table 2 also provides the estimated h-index (h_{est}) for Tsallis with $q = 1.6$, Weibull with $c = 0.4$ and Pareto with $b = 0.75$ and $a = 3$ in Eq. (4), providing $E\{h\} = 0.75(C_T/N_T)^{2/3} N_T^{1/3}$, exactly as in (Schubert and Glänzel, 2007). Thus, one can easily assess the estimation error

9

h_{meas} - h_{est}, for the provided models.[4] Fig. 3 visualizes the absolute estimation error $|h_{meas}$ - $h_{est}|$ for the above three models, as well as the Power Law citation distribution with $\alpha = 3$ (Eq. (3)); specifically, Fig. 3 provides scatter plots of the triples $\{C_T / N_T, N_T, | h_{meas} - h_{est} |\}$ that correspond to the 12 rows of Table 2. In all scatter plots, the diameter of each circle centered at $\{C_T / N_T, N_T\}$, is proportional to the absolute estimation error $| h_{meas}$ - $h_{est} |$, for each row of Table 2. It is shown that Tsallis outperforms the other three models.

Furthermore, the root-mean-squared (RMS) estimation error for Tsallis, Weibull, Power Law and Pareto becomes 2.69, 9.18, 9.76, 5.3 respectively. The data-point where Weibull and Pareto/Power Law distributions provide the highest estimation error is centered at $(C_T / N_T, N_T) = (15.18,\ 5736)$ and corresponds to last line of Table 2; if that data-point is omitted, then the RMS estimation error for Tsallis, Weibull, Power Law and Pareto becomes 2.4, 2.76, 5.32, 2.87 respectively, based on the remaining 11 rows of Table 2; in all cases, Tsallis provides the smallest estimation error, even though no special effort was made to optimize its two parameters. Fig. 3 demonstrates that:

a) the smallest root-mean-squared (RMS) h-index estimation error is given by Tsallis distribution for the studied dataset,

b) the 12 triples $\{C_T / N_T, N_T, | h_{meas} - h_{est} |\}$ can be visually deciphered only with Tsallis citation distribution, as opposed to the other three tested distributions.

This comes at no surprise given that Power Law/Paretian citation distribution better models highly-cited papers (Redner, 1998), Weibull (Stretched Exponential) better models less-cited papers (Laherrere and Sornette, 1998) and Tsallis encompasses both highly-cited and less-cited papers observed by Tsallis and de Albuquerque (2000) and recently confirmed by Burrell

---

[4] section 4 discusses a way to optimize the parameter selection for each model.

(2008). Even though the new model with Tsallis outperforms state-of-the-art, the focus of this work is to provide higher moments of the h-index and relevant scaling methodology for *any* citation distribution. Future work could provide for more accurate (than Tsallis) citation distributions.

Table 2 (and its visualization in Fig. 3) also shows that large differences on the order of 2-2.5 can be observed regarding average citations per publication, even among *related* scientific fields. For example, a scientist with $N_T = 100$ publications, working in a field with $E\{X\} \approx 10$ citations per publication (as in IEEE Transactions on Antennas & Propagation) should have h-index close to 15 on average (according to Fig. 1-right), while a scientist with the same number of publications, working in a field with $E\{X\} \approx 10$ (as in IEEE Transactions on Information Theory) should have an h-index close to 20, on average; such non-negligible difference on the order of 33.3% is observed within *related scientific fields*. Therefore, ranking according to h-index, or any other index based on citation measurements, requires extensive context-of-research definition that could be inferred from type/category of publication venues each researcher presents the bulk of her/his work; such categorization requires additional effort during measurements and automated machine procedures could be of tremendous help.

Finally, it is observed from Table 2 that the number of self-citations ($C_T - C_0$) is nearly half of the total number of citations (C_T). This observation needs further investigation and it is coherent with previous studies showing that self-citations considerably affect the h-index (Schreiber, 2007a, 2007b).

4. Ranking challenge

4.1 Measurements

The study was originally planned to provide *rankings* of similar-discipline departments mainly in Greece, according to their measured bibliometric indices. The collected bibliometric indicators included the total number of papers N_T and total number of citations C_T, the number of single-author N_0 and first-author N_1 publications, the number of journals published the last 3.5 years of the measurement window Δ_3, in some cases the total citations excluding self-citations C_0 and finally, the h-index. The measurement campaign was conducted online during the summer of 2007 until September of the same year, using WoS. Information regarding publication activity of each researcher was collected for all years available in WoS, which at the time included mainly journal publications. Personnel information regarding full professors of each department was collected from the departmental web sites, while the 'Distinct Author' WoS tool was often employed to disambiguate different researchers with common names; whenever there was doubt, or the researcher name was too common, the researcher was excluded from the dataset.

Up to September of 2007, bibliometric information was gathered for 588 full-professors stationed at Greek university departments, including Engineering (Electrical, Mechanical, Chemical, Civil, Urban Planning) and Science (Physics, Chemistry, Math, Computer Science & Informatics). Publishing culture, in terms of journals, is widespread in the aforementioned scientific fields, while WoS even though incomplete, includes most widely known relevant journals. Reference values from university departments abroad were also collected (such as MIT Electrical and Computer Engineering Department). For ranking purposes, we initially assumed that the incompleteness of the online repository should affect equally all relevant departments.

An obvious question arises: which citation distribution is universally most-appropriate and what parameters should be used? As already shown, the parameters depend on the citation window length and Tsallis seems to be a reasonable choice, as it encompasses both highly and less-cited publications (Tsallis and de Albuquerque, 2000) and it is able to describe many

observed informetric phenomena over the full distribution range (Burrell, 2008). Repositories such as Wos, could perhaps automate calculation of relevant citation histograms and use those histograms for the average h-index calculation according to the methods described before.

Alternatively, measurements could be collected and attempted to fit one of the two-parameter citation distributions mentioned above. The maximum likelihood method is one common approach to fit measurements to a given model. Laherrere and Sornette (1998) provide the maximum likelihood estimates for the two parameters (c, x_0) of Weibull distribution (Table 1), while Shalizi (2007) (and references therein) provides the maximum likelihood estimates for the case of Tsallis distribution (parameters q, λ in Table 1). The method employed throughout this work was simpler: the average value $E\{X\}$ of citations per publication was matched to the measured average value of the particular field of study, while the parameters c and q were set close to empirical values.

It was soon realized that the measured h-index for each researcher is highly correlated with her/his total number of citations C_T as well as number of publications N_T (measured correlations 0.88 and 0.82 respectively). The same finding was reported by van Raan (2006), who also observed strong correlation between h-index and C_T or N_T, with total citations correlating stronger than total number of publications, as in our measurements. More importantly, the study revealed significant variance in research domains, even among departments of the same discipline e.g. among physics departments or electrical engineering departments (Bletsas and Sahalos, 2008). The dependence of h-index on the citation distribution (or at minimum, on the average citations per publication per field $E\{X\}$) emphatically directs additional measurements before any ranking attempt; specifically, horizontal classification of all researchers according to their respective field of research is required, regardless of the

department they work at, additionally to the vertical measurements of standard, WoS-provided bibliometric information, such as h-index, C_T or N_T.

Such classification requires additional, non-trivial work that we did not perform, with the exception of Greek full-professors in signal processing. That subset is provided in Fig. 4, depicting the measured h-index and the measured number of citations per publication C_T / N_T for 19 full professors, working in the broad area of signal processing and stationed at Greek Universities. The median number of citations per researcher $<C_T> = 252$ and median number of publications per researcher $<N_T> = 42$ is also reported, for the specific dataset and measurement window.

4.2 H-index ranking case study

Five of them are selected for case study analysis which aims to highlight relevant challenges and provides a specific methodology. Table 3 offers their bibliometric information, including number of single-author journal publications N_0, number of first/single-author journal publications N_1 and number of journal publications Δ_3 the last 3.5 years (2004-Sep. 2007), with measurements up to September 2007. All 5 people are stationed in departments of electrical engineering, computer science or informatics. The national $E\{X\}$ of signal processing full-professors in Greece was estimated to $E\{X\} \approx 252/42 = 6$ citations per journal publication, as of 2007. Furthermore, $N_T = 5736$ publications appeared in IEEE Transactions on Signal Processing within 1993-2007, producing $C_T = 87070$ citations until 2007 (last row of Table 2). Therefore, this 15-year window provided an estimated average of $87070/5736 = 15.18$ citations per publication for signal processing (nearly tripled than the Greek average for full-rank professors). This is not surprising, for two reasons: a) the above calculation of $E\{X\}$ considers the set of publications appeared only at a top journal; the global average (including the Greek) should

include all papers (and their corresponding citations) published in all relevant signal processing journals, including those which are not that popular; consecutively, the global ratio of total citations per publication ratio should be significantly smaller than 15.18 for a 15-year window, b) the above calculation of $E\{X\} \approx 252/42 = 6$ assumes a 15-year publication interval, which is only approximately the research age of a full-professor.

If only h-index is used, researcher 19 will be ranked first and researcher 10 will be ranked last; interestingly, exactly the same order is preserved if all these 5 scientists are ranked according to number of journal papers N_T. On the contrary, if only number of citations per publication per researcher C_T / N_T is utilized, researcher 10 will be ranked second and researcher 19 will be ranked last. It is emphasized that such 5-professor dataset is chosen almost randomly: researchers 5 and 19 are the two researchers with the highest measured h-index among the sample of signal processing professors, while the other three were selected randomly. Assuming Tsallis citation distribution with $q = 1.6$ and average citations per publication $E\{X\} = 6$, as calculated above, the expected value $E\{h\} \equiv h_{est}$ of h-index and its standard deviation σ_h are computed for all researchers, according to the second h-index calculation method described above (first shaded column of Table 3).

Table 3 shows how many standard deviations the measured h-index is above or below the estimated average value for each observed N_T per researcher (second shaded column of Table 3). Such scaling of the h-index, takes into account field of research (i.e. ``signal processing''), geographical context (i.e. ``$E\{X\} = 6$ in Greek Univ. for signal processing as of 2007'') and, to some extent, research age (i.e. ``full-rank professors only''); it shows that researcher 5 is significantly more acknowledged than researcher 19, proportionally to her/his number of publications ($+6.85\,\sigma_h$ versus $+5.23\,\sigma_h$ respectively), even though she/he has smaller h-index,

while researcher 10, who has the smallest h-index among the five, comes third ($+2.66\sigma_h$). One could argue that researchers 10 and 16, who have roughly obtained the same h-index (7 and 8 respectively), are equally acknowledged; however, researcher 16 has achieved such figure with almost doubled number of total publications, compared to researcher 10. The fact that additional publications of researcher 16 have not attracted attention (and additional citations compared to 10) is a useful piece of information that should not be neglected. That is why researchers 10 and 16 cannot be characterized as equally acknowledged; the above is another example of why appropriate scaling of the h-index is needed before meaningful relevant rankings. Similar arguments can be made for researchers 5 and 19, who also offer similar h-indices (close to 19), while researcher 5 has achieved such number with almost 1/3 of researcher's 19 publications. Could they be characterized as ``equally- acknowledged''? The authors of this work believe that they cannot.

Furthermore, either h-index or citations per publication per researcher C_T / N_T provide inverse ranking results, which are both incoherent with the scaling method presented in this work; the latter is based on empirical citation distribution models found in practice, as well as careful calculation of the h-index probability distribution. Table 3 also provides the measured ratio h / N_T, which is an estimate of $\Pr\{X > h\}$, i.e. the probability a publication receives more than h citations (as explained in Section 2.1). Thus, comparing researchers with the same h-index becomes possible in terms of citation rate, without any knowledge requirements for the underlying citation distribution. As already mentioned, researchers 10 and 16 (as well as researchers 5 and 19) have similar measured h-index. Their measured ratio h / N_T provides ranking results equivalent to that from scaling (with Tsallis citation distribution), as presented above. It is noted however that the scaling method of this work is vastly different than scaling h-

index with total number of publications (ratio h / N_T). The statistical uncertainty of such ratio was recently studied in (Lehmann, Jackson and Lautrup, 2008).

From Table 3, it can be seen that all researchers are active the last 3.5 years, with Δ_3 following the personal publication rate of each researcher e.g. researcher 10 chooses to publish less papers per year, while researcher 19 chooses the opposite, and presents an impressive Δ_3 record (38 papers per 3.5 years or 5.43 publications per six months). Compared to 19, researcher 5 has published three times less papers, which have proportionally produced more visibility, while her/his N_0, N_0 / N_T, N_1 / N_T indices are also increased. All researchers provide relevantly high N_1 as well as N_1 / N_T.

Interestingly, 3 out of 5 researchers above have received best-paper awards from the IEEE Signal Processing Society; such international awards are annual, highly selective and prestigious. Judging from the N_T or h-index, one might guess that most likely, researcher 19 (highest h-index and/or N_T) has received such award while researcher 10 (smallest h-index and smallest N_T) is among the unlucky. Reality points to exactly the opposite (researcher 10 has been awarded and not researcher 19); remember that according to our analysis, researcher 10 is highly-acknowledged, despite her/his small (compared to her/his colleagues) h-index or N_T.

Furthermore, the average number of citations per publication can be set to $E\{X\} = 15.18$ (instead of $E\{X\} = 6$), corresponding to highly selective publications worldwide, as explained above. In that case, all five researchers provide scaled h-indices, within one standard deviation from the average approximately, with the exception of researcher 5, who is well above the average. Therefore, one could safely argue that all considered researchers above are visible among the people who publish in highly-selective, international signal processing journals. This observation highlights the obvious: bibliometric rankings are inherently comparison studies and

thus, they are highly dependent on the selected baseline reference (e.g. ``Greece'' vs ``worldwide'' vs ``top publication venues worldwide'').

From the above, it is clear that utilizing C_T / N_T alone or h-index alone, to evaluate visibility of researchers provides conflicting results or equivalently, such visibility ranking is subjective. The relationship between h-index, N_T and E{X} is non-linear and thus, scaling is required, based on the calculated h-index probability distribution. Interestingly, the minimum information required for h-index scaling and comparisons, is the average number of citations per publication E{X} *for the particular scientific field* (apart from measured h-index per researcher and total number of papers N_T). This might be surprising, given that the h-index was originally invented as a more representative yardstick compared to the average number of citations per publication *for a particular researcher*.

This observation could explain the statistical uncertainty of the h-index, reported in (Lehmann, Jackson, and Lautrup, 2005 and 2006), where it was shown through simple Bayesian inference that *un-scaled* h-index ranking is as relevant to scientific activity/visibility as ranking according to name-alphabetical order, i.e. it is irrelevant; it was suggested that the average citations per publication per author should be preferred.

5. Conclusion

This work showed that both number of publications N_T, as well as average number of citations per publication per scientific field E{X}, is needed for meaningful h-index rankings, apart from the measured h-index of each researcher. Furthermore, we provided a scaling and ranking methodology based on the h-index expected value and standard deviation, which can be

calculated through the h-index probability distribution (several examples were provided, assuming knowledge of N_T and E$\{X\}$).

This work complements previous reports on h-based rankings, where it was shown that recruitment decisions are correlated with metrics of citations per publication at the Hirsch-core, i.e. highly cited papers and their corresponding number of citations (Bornmann, Mutz and Daniel, 2008). That finding implied that careful h-index scaling is required in relevant ranking campaigns, providing additional motivation for this work.

Future work should focus on multi-parametric methods that alleviate poor practices, such as citing papers that have not been actually read (Simkin and Roychowdhury, 2006), introductory citations without real significance to the presented work, gratuitous authorship, citation coalitions and disproportional citations to widely-known researchers (``Mathew effect'') (Merton, 1968). Furthermore, self-citations in top international journals could reach roughly 50% of the total citations (as observed in this work) and thus, bibliometric indicators should be modified accordingly. Relevant corrections of the h-index regarding self-citations (e.g. see work by Schreiber 2007a, 2007b) or multi-authorship (e.g. see work by Chai, Hua, Rousseau, Wan, 2008, Schreiber 2008, Egghe 2008) have already started to appear in the literature.

Acknowledgments

References

Times Higher Education Supplement, (2008). London, UK, available at http://www.timeshighereducation.co.uk/, last access, December 2008.

Bletsas, A. and Sahalos, J., (2008). Probability distribution, scaling and ranking challenges behind the h-index. *RadioCommunications Laboratory (RCL) Technical Report*, Department of Physics, Aristotle University of Thessaloniki.

Bornmann, L., Mutz, R., and Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. *Jour. Americ. Soc. Inf. Scienc. Tech.*, **59 (5)**, 830--837.

Burrell, Q. L. (2007). Hirsch's h-index: a stochastic model. *Journal of Informetrics*, **1 (1)**, 16--25.

Burrell, Q. L. (2008). Extending Lotkaian Informetrics. *Inf. Process. Manage.,* **44 (5)**, 1794--1807.

Chai, J. C., Hua, P. H., Rousseau, R. and Wan J. K. (2008). The adapted pure h-index. *Proceedings of Fourth International Conference on Webometrics, Informetrics and Scientometrics (WIS) 2008*, Berlin.

Egghe, L. and Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, **69 (1)**, 121--129.

Egghe, L. (2007). Dynamic h-index: the Hirsch index in function of time. *Jour. Americ. Soc. Inf. Scienc. Tech.*, **58 (3)**, 452--454.

Egghe, L. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. *Jour. Americ. Soc. Inf. Scienc. Tech.*, **59 (10)**, 1608--1616.

Guns, R. and Rousseau, R. (2009). Real and rational variants of the h-index and the g-index. *Journal of Informetrics*, **3 (1)**, 64--71.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc. Nat. Acad. Scienc.*, **102 (46)**, 16 569--16 572.

Hirsch, J. E. (2007). Does the h index have predictive power? *Proc. Nat. Acad. Scienc.*, **104 (49)**, 19 193--19 198.

Iglesias, J. E. and Pecharroman, C. (2007). Scaling the h-index for different scientific isi fields. *Scientometrics*, **73 (3)**, 303--320, also available at http://arxiv.org/abs/physics/0607224.

Laherrere, J. and Sornette, D. (1998). Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *Eur. Phys. J. B*, **2 (4)**, 525--539.

Lehmann, S., Jackson, A. D. and Lautrup, B. E. (2005). Measures and mismeasures of scientific quality. *arXiv.org*, available at http://arxiv.org/abs/physics/0512238.

Lehmann, S., Jackson, A. D. and Lautrup, B. E. (2006). Measures for measures. *Nature*, **444 (21)**, 1003--1004.

Lehmann, S., Jackson, A. D. and Lautrup, B. E. (2008). A quantitative analysis of indicators of scientific performance. *Scientometrics*, **76 (2)**, 360--390.

Merton, R. K. (1968). The Mathew Effect in Science. *Science*, **159 (3810)**, 56--63.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B*, **4 (2)**, 131--134.

Ruane, F. and Tol, R. S. J. (2008). Rational (successive) h-indices: an application to economics in the Republic of Ireland. *Scientometrics*, **75 (2)**, 395--405.

Schreiber, M. (2007 a). A case study of the hirsch index for 26 non-prominent physicists. *Ann. Phys. (Leipzig)*, **16 (9)**, 640--652.

Schreiber, M. (2007 b). Self-citation corrections for the hirsch index. *EPL*, **78 (3)**, 30002 (6pp).

Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. *Jour. Americ. Soc. Inf. Scienc. Tech.*, **59 (9)**, 1513--1522.

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

Schubert, A. and Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, **1 (3)**, 179--184.

Shalizi, R. C. (2007). Maximum likelihood estimation for q-exponential (Tsallis) distributions. arXiv.org (arXiv:math/0701854v2, http://arxiv.org/abs/math.ST/0701854).

Simkin M. and Roychowdhury V. (2006). Do you sincerely want to be cited? Or: read before you cite, *Significance*, **3 (4)**, 179--181.

Tsallis, C. and de Albuquerque, M. P. (2000). Are citations of scientifc papers a case of nonextensivity? *Eur. Phys. J. B*, **13 (4)**, 777--780.

van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, **67 (3)**, 491--502.

TABLES

Table 1: Continuous Empirical Citation Distributions

(X : number of citations per publication per scientific field)

	$f(x)$	$F_c(x) = \Pr\{X > x\}$	$\mathsf{E}\{X\}$	$Y_\kappa \left(F_c(Y_\kappa) = \kappa / N_T \right)$
Power Law	$[(a-1)/x_{\min}](x/x_{\min})^{-a}$ $(x \geq x_{\min} > 0, a > 1)$	$(x/x_{\min})^{-(a-1)}$	$\dfrac{a-1}{a-2}x_{\min}, a > 2$	$x_{\min}(N_T/\kappa)^{[1/(a-1)]}$
Weibull	$(c/x)(x/x_0)^c \exp\!\left(-(x/x_0)^c\right)$ $(x > 0, x_0 > 0, c > 0)$	$\exp\!\left(-(x/x_0)^c\right)$	$x_0\,\Gamma\!\left(1+\dfrac{1}{c}\right)$	$x_0\left[\ln(N_T/\kappa)\right]^{(1/c)}$
Tsallis	$\lambda\left[1+\lambda(q-1)x\right]^{-\frac{q}{q-1}}$ $(x \geq 0, \lambda > 0, q > 1)$	$\left[1+\lambda(q-1)x\right]^{-\frac{1}{q-1}}$	$[\lambda(2-q)]^{-1}, q < 2$	$\dfrac{1}{\lambda(q-1)}\left[(N_T/\kappa)^{(q-1)}-1\right]$

Table 2: Measured and Estimated Indices of IEEE Transactions Relevant to

Wireless Communications (Citation Window: up to Dec. 2008 , Source:WoS)

	IEEE Transactions on	Pub. Year	N_T	C_T (C_0)	$\dfrac{C_T}{N_T}$	h_{meas}	h_{est} Tsallis	h_{est} Weibull	h_{est} Pareto
1	Information Theory	1993	182	4505 (2941)	24.75	**32**	**31.12**	**31.02**	**36.09**
2	Signal Processing	1993	353	5866 (3556)	16.62	**37**	**34.02**	**37.57**	**34.52**
3	Communications	1993	242	3683 (2285)	15.22	**28**	**27.37**	**29.47**	**28.70**
4	Antennas & Propagation	1993	281	3741 (2678)	13.31	**31**	**27.17**	**29.99**	**27.58**
5	Information Theory	2002	283	7005 (3106)	24.75	**38**	**38.29**	**39.95**	**41.81**
6	Signal Processing	2002	288	6988 (3142)	24.26	**40**	**38.19**	**39.99**	**41.50**
7	Communications	2002	228	5333 (2391)	23.39	**36**	**33.60**	**34.50**	**37.47**
8	Antennas & Propagation	2002	221	2136 (1365)	9.67	**20**	**20.45**	**22.72**	**20.58**
9	Signal Processing	2003	289	5349 (2392)	18.51	**37**	**33.02**	**35.48**	**34.69**
10	Communications	2003	307	5602 (2219)	18.25	**30**	**33.66**	**36.43**	**35.07**
11	Antennas & Propagation	2003	487	4678 (2671)	9.61	**29**	**28.65**	**33.70**	**26.67**
12	Signal Processing	1993-2007	5736	87070 (22185)	15.18	**98**	**102.86**	**128.47**	**82.31**

Table 3: Bibliometric indicators for case study of 5 signal processing full-professors.

(Citation Window: up to Dec. 2008 , Source:WoS.

Tsallis distribution with $\mathrm{E}\{X\} = 6$ and $q = 1.6$ has been utilized.)

ID	h_{meas}	N_T	$\dfrac{C_T}{N_T}$	$\dfrac{h_{meas}}{N_T}$	$N_0\left(\dfrac{N_0}{N_T}\right)$	$N_1\left(\dfrac{N_1}{N_T}\right)$	Δ_3	$\mathrm{E}\{h\}, \sigma_h$	$\dfrac{h_{meas} - \mathrm{E}\{h\}}{\sigma_h}$
5	18	59	22.13	0.3	11 (0.19)	26 (0.44)	13	8.4, 1.4	+6.85
10	7	16	15.75	0.44	1 (0.06)	8 (0.50)	2	4.34, 1.00	+2.66
12	10	45	7.86	0.22	4 (0.09)	20 (0.44)	16	7.38, 1.3	+2.01
16	8	30	9.13	0.27	2 (0.07)	10 (0.33)	8	6.05, 1.13	+1.72
19	20	160	7.69	0.12	5 (0.03)	35 (0.22)	38	13.41, 1.26	+5.23

FIGURE CAPTIONS

Figure 1: Average h-index as a function of average citations per publication ($\mathsf{E}\{X\}$) and total number of papers (N_T) for two cases of Tsallis Distribution. N_T increases in steps of 40 publications.

Figure 2: Average h-index as a function of average citations per publication ($\mathsf{E}\{X\}$) and total number of papers (N_T) for two cases of Weibull (Stretched Exponential) Distribution. N_T increases in steps of 40 publications.

Figure 3: Each circle corresponds to a line of Table 2. Absolute h-index estimation error is proportional to the radius of each circle, centered at $(C_T/N_T, N_T)$. Four empirical models of citation distribution are exploited for twelve datasets relevant to electrical engineering journals (12 lines of Table 2). Root-mean-squared (RMS) estimation error is also reported; Tsallis citation distribution provides the smallest h-index estimation error.

Figure 4: Measured h-index and number of citations per journal publication (C_T/N_T) for 19 full professors of signal processing. Median number of citations C_T and publications N_T per researcher are also depicted. Measurement Window: up to September 2007. Source: WoS.
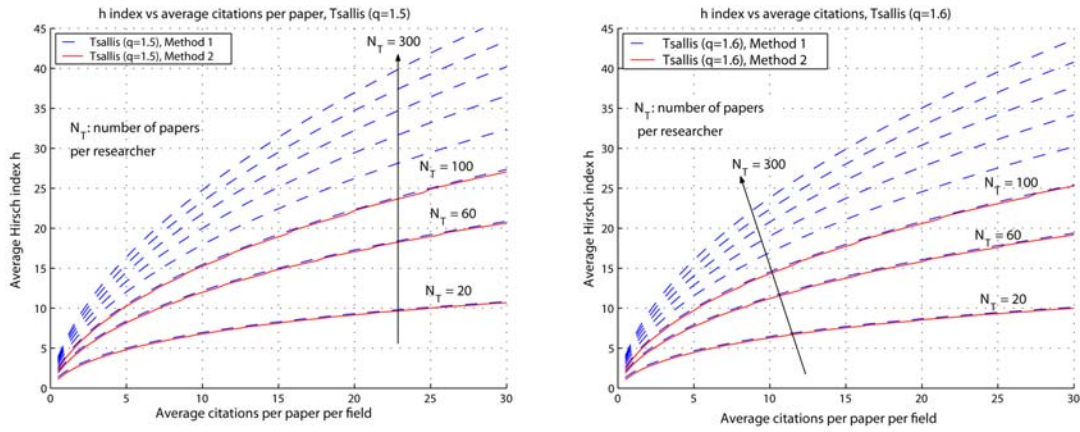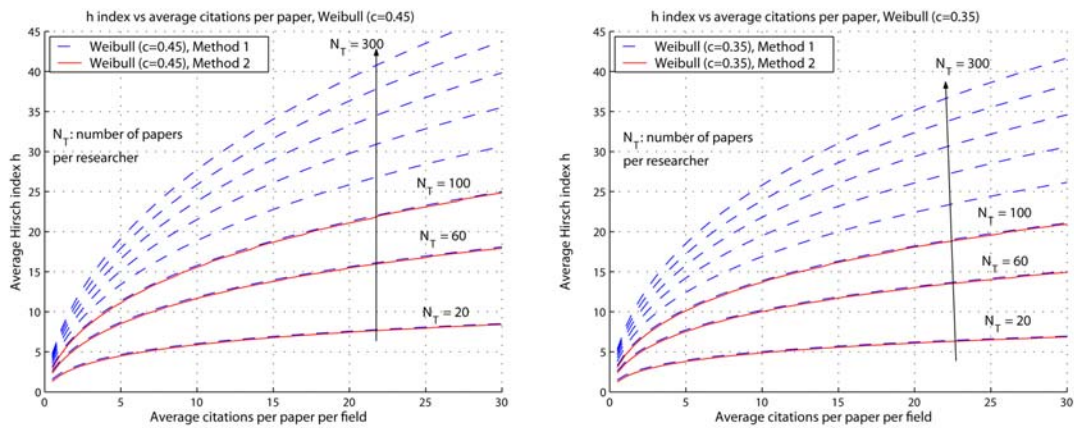
Figure 1: Average h-index as a function of average citations per publication ($\mathsf{E}\{X\}$) and total number of papers (N_T) for two cases of Tsallis Distribution. N_T increases in steps of 40 publications.

Figure 2: Average h-index as a function of average citations per publication ($\mathsf{E}\{X\}$) and total number of papers (N_T) for two cases of Weibull (Stretched Exponential) Distribution. N_T increases in steps of 40 publications.
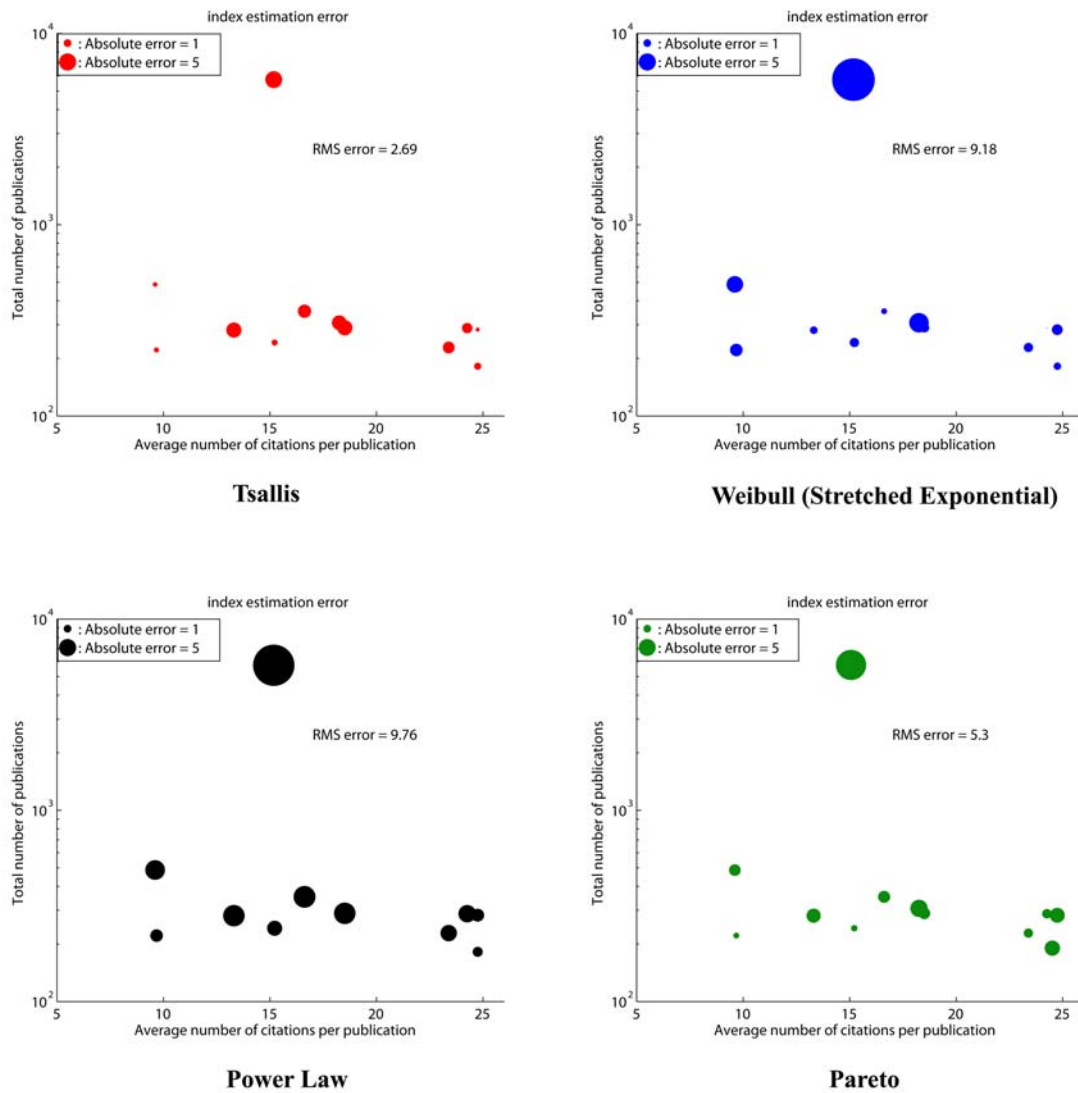
Figure 3: Each circle corresponds to a line of Table 2. Absolute h-index estimation error is proportional to the radius of each circle, centered at $(C_T/N_T, N_T)$. Four empirical models of citation distribution are exploited for 12 datasets relevant to electrical engineering journals. Root-mean-squared (RMS) estimation error is also reported; Tsallis citation distribution provides the smallest h-index estimation error.
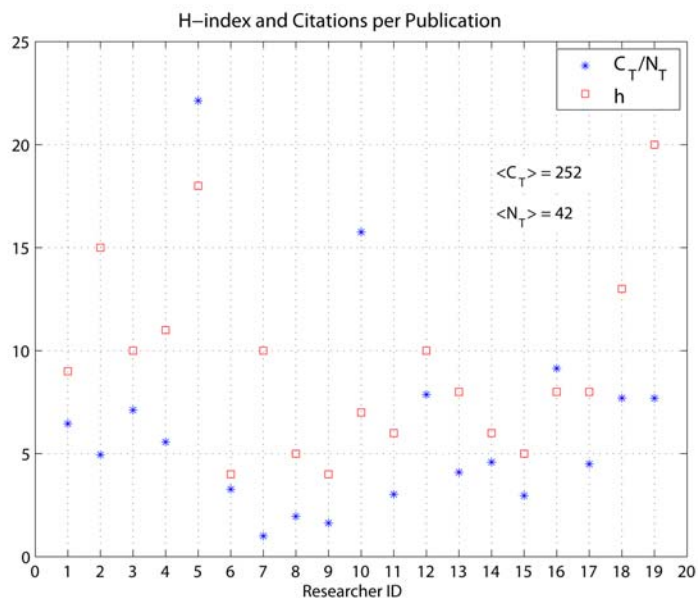
Figure 4: Measured h-index and number of citations per journal publication (C_T/N_T) for 19 full professors of signal processing. Median number of citations C_T and publications N_T per researcher are also depicted. Measurement Window: up to September 2007. Source: WoS.