Arch. Immunol. Ther. Exp., 2009, **57**, 85–90
PL ISSN 0004-069X
DOI 10.1007/s00005-009-0014-0

On the fairness of using relative indicators for comparing citation performance in different disciplines

Claudio Castellano[1] and Filippo Radicchi[2]

[1]SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Universita' di Roma, Roma, Italy
[2]Complex Networks Lagrange Laboratory, Institute for Scientific Interchange Foundation, Torino, Italy

Received: 2009.03.04, **Accepted:** 2009.03.11, **Published online:** 2009.03.31

Abstract

Relative indicators are commonly used to remove biases due to different citation practices in various scientific fields. Here we extend our recent investigation on the viability of the use of relative indicators for comparing article impact in different disciplines. We consider citation distributions for papers published in 14 of the 172 disciplines categorized by the Journal Citation Reports. The distribution of the number of citations received by publications in a certain discipline divided by the average number for the discipline is a universal function. Based on it, we compute the relative number of citations needed to be among the q percent most-cited publications in a discipline. The effect of finite samples is also discussed. The average number of citations is shown to be strongly correlated with the impact factor, but fluctuations are quite large. A similar universal distribution is found (with exceptions) when citation distributions restricted to papers published in a single journal are considered.

Key words: citation analysis, relative indicators, impact factor, universality.

Abbreviations: IF – impact factor, WoS – Web of Science, Equation – Eq.

Corresponding author: Claudio Castellano, SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Universita' di Roma, P.le A. Moro 2, 00185 Roma, Italy, tel.: +39 06 49937511, fax: +39 06 49937440, e-mail: claudio.castellano@roma1.infn.it

INTRODUCTION

The use of citation counts is a widespread practice for evaluating scientists, projects, research groups, departments, etc., up to nations (Egghe 2006; Hirsch 2005; Hirsch 2007; King 2004; Kinney 2007). While the debate on the appropriateness to use such quantitative tools is still open, the use of numerical indicators derived from citation patterns is becoming more diffuse. Recently it has even reached an environment traditionally hostile to it as recruitment procedures in Italy: the national board of universities (Consiglio Universitario Nazionale) has introduced for the first time a list of minimal quantitative requirements to be fulfilled by candidates for academic positions (http://www.cun.it/media/100062/indicatori_completo.pdf).

The identification of suitable measures of scientific performance becomes therefore more and more important. One of the major problems in this respect is that there are many factors that affect the number of citations received by an article and that are not related to its scientific quality or impact. Among these factors are, for example, the year of publication, the language, and the scientific discipline (Bornmann and Daniel 2008). Their influence should be factored out for a fair quantitative comparison among publications.

One possible way to do this is to consider relative indicators, i.e. to rescale the citation numbers by some suitable average value computed over the whole set of publications considered. This method has been used for decades with remarkable success (Iglesias and Pecharroman 2007; Schubert and Braun 1986; Schubert and Braun 1996; Vinkler 1996; Vinkler 2003). Recently, an empirical study of the full citation distributions strongly validated the use of relative citation indicators for comparing articles in different scientific disciplines or published in different years (Radicchi et al. 2008). It turns out that the distributions of rescaled citations restricted to single disciplines or to specific years are the same, so that the rescaled indicator weighs in a balanced way the impact of papers across years and fields.

In this paper we pursue further the analysis of the results of Radicchi et al. (Radicchi et al. 2008) and investigate the possibility of generalizing them. In particular we derive the values of the rescaled citation count $c_f(q)$ that correspond to the top q% most-cited papers and take into account fluctuations due to the finite number of papers published. In addition, we study the relationship between the rescaling factor c_0 and the impact factor (IF). Finally, we test the presence of universality also in the distribution of citations restricted to single journals.

MATERIALS AND METHODS

Our empirical analysis is based on data from Thomson Scientific's Web of Science database (WoS, www.isiknowledge.com), in which the number of citations is counted as the total number of times an article appears as a reference in a more recent publication. Scientific journals are divided into 172 categories, from "Acoustics" to "Zoology". We consider a subset of 14 of them. Within a single category, a list of journals is provided. We consider articles published in each of these journals to be part of the category. Notice that the division into categories is not mutually exclusive; for example, Physical Review D belongs both to the "Astronomy and Astrophysics" and to the "Physics, particles and fields" categories. For consistency, among all records contained in the database we consider only those classified as "article" and "letter", thus excluding reviews, editorials, comments, and other published material likely to have an uncommon citation pattern. A list of the categories and years considered, with the relevant parameters that characterize them, is reported in Table 1.

Our calculations neglect uncited articles; we have verified, however, that their inclusion just produces a small shift in c_0 which does not affect the results of our analysis. In the plots of the citation distributions, data have been grouped in bins of exponentially growing size, so that they are equally spaced along a logarithmic axis. For each bin we count the number of articles with citation count within the bin and divide by the number of all potential values for the citation count that fall in the bin (i.e. all integers). This holds as well for the distribution of the normalized citation count c_f, as the latter is just determined by dividing the citation count by the constant c_0, so it is a discrete variable just like the original citation count. The resulting ratios obtained for each bin are finally divided by the total number of articles considered, so that the histograms are normalized to 1.

Table 1. List of all subject categories (scientific disciplines) considered in this paper. For each of them we report, from left to right: the index associated with the discipline, the name of the discipline, the year of publications, the number of papers published in that year in all journals within a subject category, the average number of citations per paper, and the maximum value of citations received by a single publication. Subject categories are defined by Journal Citation Reports.

Index	Subject category	Year	N_p	c_0	c_{max}
1	Agricultural Economics & Policy	1999	266	6.88	42
2	Allergy	1999	1530	17.39	271
3	Anesthesiology	1999	3472	13.25	282
4	Astronomy & Astrophysics	1999	7399	23.77	1028
5	Biology	1999	3400	14.6	413
6	Computer Science, Cybernetics	1999	704	8.49	100
7	Developmental Biology	1999	2982	38.67	520
8	Engineering, Aerospace	1999	1070	5.65	95
9	Hematology	1990	4423	41.05	1424
10	Hematology	1999	6920	30.61	966
11	Hematology	2004	8695	15.66	1014
12	Mathematics	1999	8440	5.97	191
13	Microbiology	1999	9761	21.54	803
14	Neuroimaging	1990	444	25.26	518
15	Neuroimaging	1999	1073	23.16	463
16	Neuroimaging	2004	1395	12.68	132
17	Physics, Nuclear	1990	3670	13.75	387
18	Physics, Nuclear	1999	3965	10.92	434
19	Physics, Nuclear	2004	4164	6.94	218
20	Tropical Medicine	1999	1038	12.35	126

Our measurements are affected by fluctuations due to finite size effects. In the particular case of order sequences of random numbers, we take advantage of known formulas of order statistics (David and Nagaraja 2003). Given a sorted sequence (in ascending order) of n random numbers extracted from a probability distribution function f(x), the probability that the k-th element of the sequence equals x is

$$P_k(x) = \frac{n!}{(k-1)!\,(n-k)!}\,F(x)^{k-1}\,[1-F(x)]^{n-k}\,f(x) \quad (1)$$

where F(x) is the cumulative distribution function of f(x). Therefore the probability that x is the threshold for the top q percentile is given by the formula above by setting k equal to the integer closest to the quantity n(1–q/100)+1. Using Equation (1), the average value and the standard deviation of x are readily calculated.

The IF (Garfield 1979) is a bibliometric indicator frequently used as a proxy for the relevance of a journal. The IF of a journal is calculated based on a two-year period and corresponds to the average number of citations in a year given to those papers in a journal that were published during the two preceding years. For example, if one wants to calculate the IF of a journal for the year 2001, he counts the number of papers published in that journal during 1999 and 2000 (namely, N) and the number of citations received by them from papers published (everywhere) in 2001 (namely, C). The IF of the journal will be given by the ratio C/N. Since we are interested in calculating the IF not for journals, but for disciplines, we first take the IF for all journals within a discipline and then calculate the IF of this discipline as the weighted average of the IFs of all journals belonging to it. In particular, we calculate the IF of a given discipline for the year 2001 as

$$\mathrm{IF} = \frac{\sum_j N_j\,\mathrm{IF}_j}{\sum_j N_j} \quad (2)$$

where the sum runs over all journals j in the discipline, IF_j is the IF of the j-th journal in 2001, and N_j is the total number of papers published in the j-th journal in 1999.

RESULTS

The distribution of the number of citations received by an article strongly depends on the scientific discipline. This is made clear in Fig. 1, where the distribution P(c) vs. c is plotted for several disciplines spanning all main areas of science. It turns out that the same number of citations may be a sign of rather poor impact in some discipline, but indicates a very successful paper in another. However, it is visually obvious that there is a strong correlation between the width of the curve for a discipline and the value c_0 of the average number of citations per paper in the same discipline. This suggests that the use of the rescaled indicator $c_f=c/c_0$ may reduce or eliminate the bias in favor of disciplines for which the average number of citations per paper is higher. Fig. 2 shows evidence that such an attempt works: the distribution of c_f is the same for all the disciplines considered. Fig. 2 allows us then to write the distribution of citations restricted to publications about a certain discipline as

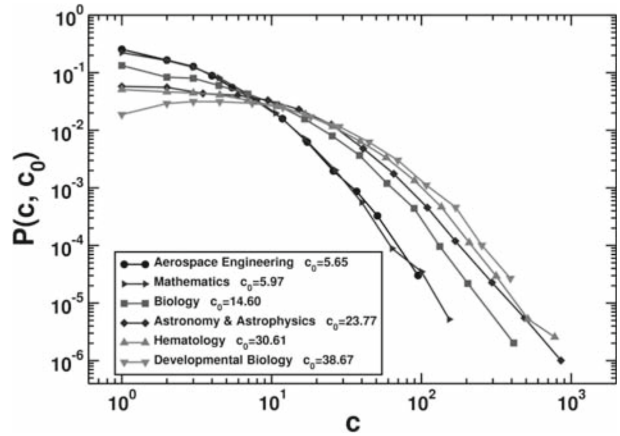

Fig. 1. Probability distribution $P(c,c_0)$ of papers published in 1999 and having received c citations. Different colors and symbols correspond to different scientific disciplines. The average number of citations per paper (c_0) depends on the specific discipline considered.

$$P(c, c_0) = \frac{1}{c_0}\,Q(c_f) \quad (3)$$

where $Q(c_f)$ is a universal curve independent of the discipline. The universal curve has a maximum and a tail that decays faster than a power law and more slowly than an exponential. A numerical fit with a lognormal shape

$$Q(c_f) = \frac{1}{c_f\,\sigma\,\sqrt{2\pi}}\,\exp\left\{-\frac{[\ln(c_f)+\sigma^2/2]^2}{2\sigma^2}\right\} \quad (4)$$

gives quite good agreement with the empirical results over the whole range of values of c_f, with a parameter $\sigma=1.14$.

Using the analytical expression of the universal curve it is possible to derive some additional useful results. Integrating the function Q from $c_f(q)$ to infinity one obtains the percentage q of all papers that have a relative impact larger than $c_f(q)$

$$\frac{q}{100} = \int_{c_f(q)}^{\infty} dc'_f\,Q(c'_f) = \frac{1}{2}\,[1-\mathrm{erf}(v)] \quad (5)$$

where

$$v = \frac{\ln(c_f)+\sigma^2/2}{2\,\sigma^2}$$

and erf(x) stands for the error function. Inverting this relation one obtains

$$c_f(q) = \exp\left[2\,\sigma^2\,\mathrm{erf}^{-1}(1-2\,q/100) - \sigma^2/2\right] \quad (6)$$

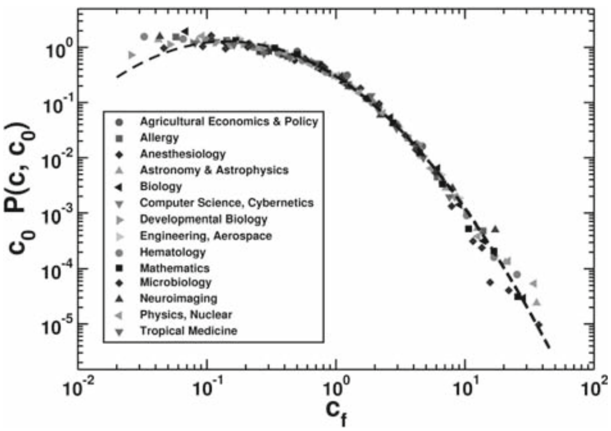

Fig. 2. Rescaled probability distribution c_0 P(c,c_0) as a function of the relative indicator c_f =c/c_0 for papers published in 1999 and belonging to the disciplines listed in Table 1. The rescaled probability distributions show a clear universal scaling, since all curves collapse into the same distribution function. The dashed line corresponds to a lognormal distribution [Eq. (4)] with σ=1.14.

which gives the minimum value $c_f(q)$ for a paper to be in the top q percent most-cited articles of its discipline. This curve is plotted in Fig. 3. It turns out then that only 28.4% of the papers have a value of c_f larger than 1. To belong to the top 1 percentile in its own discipline, a paper must be cited at least 7.4 times more than the average in the category. Some special values of the function $c_f(q)$ [Eq. (6)] and of its inverse $q(c_f)$ are reported for reference in Table 2.

When considering practical cases, however, it is important to take into account the fact that the number of articles published in a year and in a discipline is finite. This has the consequence that even if we assume that Eq. (4) perfectly describes the citation statistics, the figures presented in Table 2 are only expected values and in practice there will be fluctuations around them. To characterize such fluctuations, we present in Fig. 4 the value of $c_f(q)$ determined empirically for the various disciplines and years we analyze compared with the predictions of Eq. (6). It turns out, as expected, that for q=50, 10, and 5%, fluctuations are quite narrowly dispersed around the theoretical prediction, confirming the validity of the numerical fit to the universal function $Q(c_f)$. For small values of q (1%), fluctuations are rather large. It is possible to determine the expected amplitude of such fluctuations due to the finiteness of the sample of publications (see Material and Methods section). The error bars computed in this way make the empirical data compatible with the theoretical values in almost all

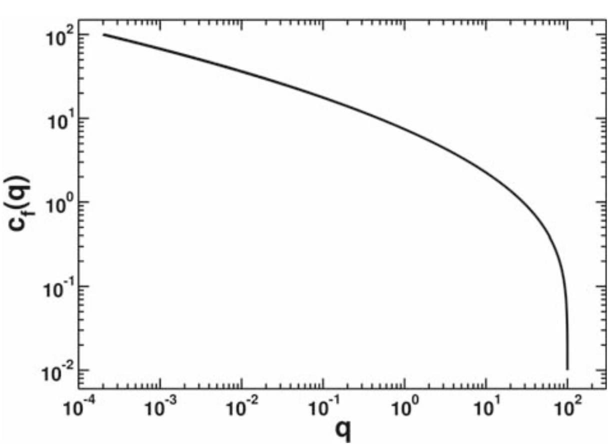

Fig. 3. Value of the rescaled indicator c_f needed to be in the top q percent most-cited papers [Eq. (6)].

Table 2. Left: threshold values of the number of rescaled citations c_f for a paper to be among the top q percent most-cited values in its discipline, calculated according to Eq. (6). Right: value of q as a function of c_f [Eq. (5)]

q	$c_f(q)$	c_f	$q(c_f)$
0.01	36.247	0.1	92.6389
0.1	17.698	0.5	51.5093
1	7.408	1	28.4309
5	3.406	2	11.9395
10	2.251	5	2.37588
20	1.363	10	0.480454
50	0.523	20	0.0693077

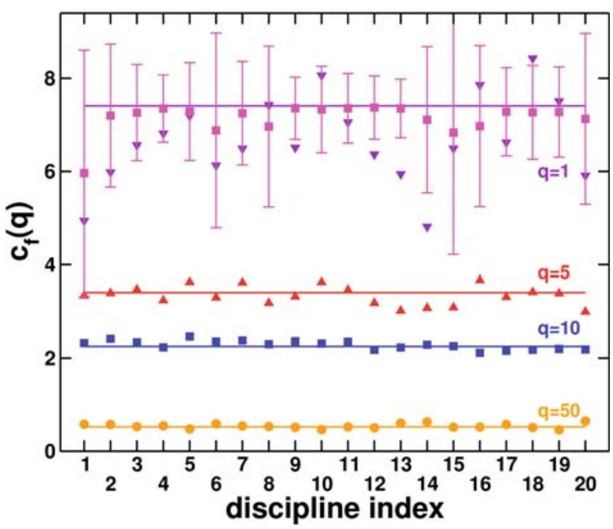

Fig. 4. Threshold values [$c_f(q)$] of the relative indicator corresponding to the top q percent of papers published in each of the years and scientific disciplines listed in Table 1. We consider q=50 (orange circles), 10 (blue squares), 5 (red up-triangles), and 1 (violet down-triangles). The horizontal solid lines correspond to our theoretical predictions (see Table 2), valid in the infinite size limit. For q=50, 10, and 5, the empirical values of $c_f(q)$ are well fitted by our theoretical predictions. For q=1, finite size effects become relevant. We compute in this case the theoretical expectations for finite sample sizes of the average values (magenta squares) and the standard deviations of $c_f(q)$. For almost all disciplines and years the empirical and theoretical values of $c_f(q)$ differ by less than two standard deviations (represented by the error bars).

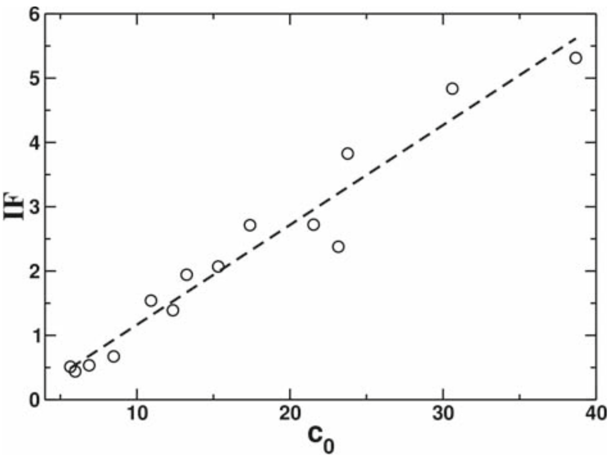

Fig. 5. 2001 IF for single disciplines [as defined in Eq. (2)] as a function of the average number of citations per paper c_0 in 1999. Each point corresponds to one of the scientific disciplines listed in Table 1. The figure shows a clear linear correlation between IF and c_0 in all cases (all points are close to the dashed line), but also that appreciable deviations from the linear behavior occur.

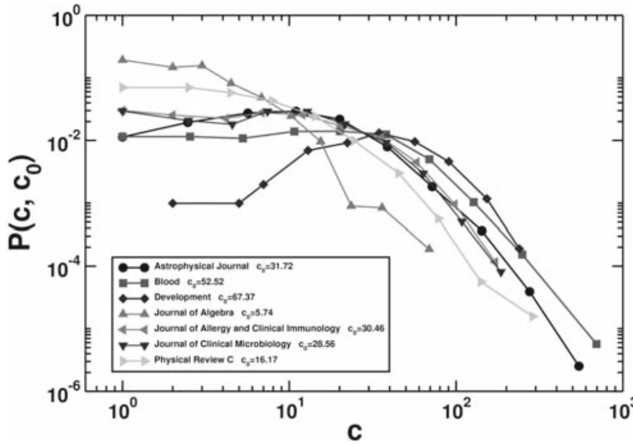

Fig. 6. Probability distribution $P(c,c_0)$ of papers published in 1999 and having received c citations. Different colors and symbols correspond to different scientific journals. The average number of citations per paper (c_0) depends on the specific journal considered and seems to be relevant for the shape of $P(c,c_0)$.

cases. Only a small systematic deviation from the value presented in Table 2 can be spotted, indicating that the quality of the numerical fit is slightly reduced for such large values of c_f.

Another relevant question posed by the universality uncovered in Radicchi et al. (Radicchi et al. 2008) (Fig. 2) is whether the quantity c_0 can be replaced by the IF, which is an indicator much more easily available. To test this hypothesis we determined the global 2001 IF for the disciplines considered (see the definition in the Materials and Methods section) and plotted it versus the 1999 value of c_0 (Fig. 5). It is evident that there is on average a linear correlation between the two indicators. However, it is also clear that quite large fluctuations occur. Hence rescaling the number of citations with IF instead of with c_0 would not be equally efficient in removing the bias.

Finally, one may wonder whether the universality property of the distributions of rescaled citations can also be found when single journals instead of disciplines are considered. To investigate this issue we computed the distribution of citations for papers appearing in journals (in the disciplines considered) that published more than 300 articles in 1999. They are reported in Fig. 6 and exhibit a large heterogeneity [former studies have suggested the possibility to fit these distributions with log-normal functions (Stringer et al. 2008)]. Also in this case we rescale (Fig. 7) with the factor c_0 (now the average number of citations obtained by papers published in the journal). The results show a large degree of universality, with all distributions except one collapsing onto a single curve. However, the behavior exhibited by the journal "Development" is remarkably different. These results indicate that some universal behavior may also be present at this level, but further investigations are needed to understand the extent and limits of this property.

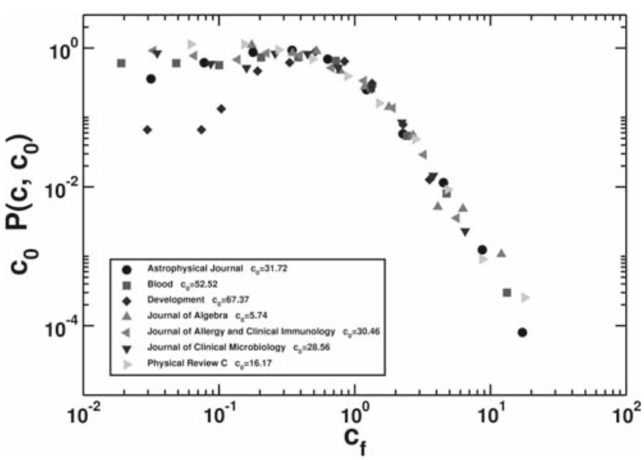

Fig. 7. Rescaled probability distribution c_0 $P(c,c_0)$ as a function of the relative indicator $c_f=c/c_0$ for all journals considered in Fig. 6. The rescaled distributions have a similar shape in the majority of the cases, but deviations from the universal behavior are evident.

DISCUSSION

In this paper we presented some results that further clarify the picture emerging from Radicchi et al. (Radicchi et al. 2008) but, on the other hand, indicate that additional work is needed in this direction. First we determined explicitly the function that gives the minimum value of the rescaled indicator $c_f(q)$ for a paper to be in the top q percentile of the universal distribution. In this way it can be immediately determined in which percentile a paper with a given value of c_f falls and, vice versa, what the minimum value of c_f must be to be in the top q% most-cited papers. We have also considered the effect on these values of the finite number of articles that are published in a discipline in a year. It turns out that the fluctuations induced by this finiteness are relevant only for very small values of q and that they are almost completely accounted for within the hypothesis that the relative indicator c_f is not biased. This result further confirms the validity of the analysis performed in Radicchi et al. (Radicchi et al. 2008). This makes more urgent the need to validate the hypothesis of universality for all scientific disciplines (and not only a subset of them). We expect the global picture to be valid for most scientific disciplines, but exceptions may occur as well.

The investigation of the relationship between c_f and the IF indicates that, on average, the two indicators provide similar information. However, the presence of large fluctuations makes the use of the IF as a rescaling factor inappropriate. This is a problem because while the IF is tabulated and readily available, c_0 is generally not known. From this point of view, the availability of the value of c_0 over the different years for all scientific disciplines categorized by WoS would be of great help.

Finally, an interesting question to be pursued further is whether the universality exhibited when papers of different disciplines or different years are considered extends to other cases. In this study we computed the rescaled distributions of citations restricted to articles published in a single journal, considering only journals with a sufficiently large number of published papers per year. The analysis reveals that some form of universality is present, but also that exceptions occur. Further work is needed to elucidate the scope and origin of this regularity.

Acknowledgment: We acknowledge Santo Fortunato for participation in the early stages of this project. We thank Simon Thorpe for useful suggestions.

REFERENCES

Bornmann L, Daniel H-D (2008) What do citation counts measure? A review of studies on citing behavior. J Doc 64: 45–80

David HA, Nagaraja HN (2003) Order Statistics, 3rd edn. Wiley, Hoboken, NJ

Egghe L (2006) Theory and practise of the g-index. Scientometrics 69:131–152

Garfield E (1979) Citation Indexing. Its Theory and Applications in Science, Technology, and Humanities. Wiley, New York

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 102:16569–16572

Hirsch JE (2007) Does the h index have predictive power? Proc Natl Acad Sci USA 104:19193-19198

Iglesias JE, Pecharroman C (2007) Scaling the h-index for different scientific ISI fields. Scientometrics 73:303–320

King DA (2004) The scientific impact of nations. Nature 430:311–316

Kinney AL (2007) National scientific facilities and their science impact on nonbiomedical research. Proc Natl Acad Sci USA 104:17943–17947

Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. Proc Natl Acad Sci USA 105:17268–17272

Schubert A, Braun T (1986) Relative indicators and relational charts for comparative-assessment of publication output and citation impact. Scientometrics 9:281–291

Schubert A, Braun T (1996) Cross-field normalization of scientometric indicators. Scientometrics 36:11–324

Stringer MJ, Sales-Pardo M, Nunes Amaral LA (2008) Effectiveness of journal ranking schemes as a tool for locating information. PLoS ONE 3:e1683

Vinkler P (1996) Model for quantitative selection of relative scientometric impact indicators. Scientometrics 36:223–236

Vinkler P (2003) Relations of relative scientometric indicators. Scientometrics 58:687–694