Rescaling citations of publications in physics

Filippo Radicchi
Howard Hughes Medical Institute (HHMI), Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA

Claudio Castellano
Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy
Dipartimento di Fisica, "Sapienza" Università di Roma, Piazzale A. Moro 2, I-00185 Roma, Italy
(Received 23 December 2010; published 22 April 2011)

We analyze the citation distributions of all papers published in Physical Review journals between 1985 and 2009. The average number of citations received by papers published in a given year and in a given field is computed. Large variations are found, showing that it is not fair to compare citation numbers across fields and years. However, when a rescaling procedure by the average is used, it is possible to compare impartially articles across years and fields. We make the rescaling factors available for use by the readers. We also show that rescaling citation numbers by the number of publication authors has strong effects and should therefore be taken into account when assessing the bibliometric performance of researchers.

DOI: 10.1103/PhysRevE.83.046116 PACS number(s): 01.30.−y, 01.40.G−, 01.78.+p

I. INTRODUCTION

Despite its many shortcomings, citation analysis is increasingly used as a quantitative tool to evaluate research performance ranging from the single publication up to the level of individual researchers, groups, departments, institutions, and countries [1–8]. The underlying idea is that the number of citations measures the impact of a publication and it is thus a proxy for the quality and importance of the scientific work described in it. This assumption is highly questionable and the empirical evidence revealing striking counterexamples to its general validity is substantial [9–11]. In principle, a careful, unbiased review by peers with no conflict of interest would be a better procedure to evaluate research quality and impact (although many distortions can also occur in this case [12]). Nevertheless, for several reasons that go much beyond the scope of this work and will not be discussed here, the use of quantitative bibliometric measures is not going to diminish in the future. It is much more realistic and effective trying to correct the inadequacies of how citation analysis is performed, rather than self-deceptively hope that citation analysis will lose importance in the future [13].

In this paper we tackle two of the more obvious problems in the use of the number of citations to measure the impact of scientific publications: (1) papers in some fields are typically more cited than papers in other fields and (2) old papers naturally tend to have more citations than more recent ones. These facts are part of the common wisdom of people working in science, yet it is common to see comparisons between the number of cites of papers dealing with completely different topics or published in different decades. However, as we will show, one should not take for granted that a paper which has been cited twice as much as another in a different field has actually had a larger impact. The most natural way to get rid of (or at least alleviate) these problems is the use of relative indicators, that is, the normalization of the number of citations by some average over a suitable reference set. The use of relative indicators has been proposed long ago [14] and used in various contexts, but often in questionable ways [15,16] and without an empirical check that it actually solves the problem. Recently we have shown that such a normalization strongly reduces biases when papers from entirely different disciplines, spanning all fields of science, are compared [17,18]. Here we present a similar analysis applied to publications of the American Physical Society (APS). With respect to Refs. [17,18] we consider a more fine-grained level of categorization since the different groups we deal with are all within the realm of physics. Fields are identified using the PACS (Physics and Astronomy Classification Scheme [19]) number scheme. We show first that relevant differences in citation patterns show up among different fields. The average number of citations in some field can be up to three times larger than for papers in another field. Analogously, we show that papers published in 1985 are cited in some fields even 15 times more than papers published in 2009. We then show that introducing a suitable relative indicator the distributions of such an indicator become essentially independent from the field or the year of publication, so that the relative indicator can be used as an unbiased measure of impact. We corroborate these results by showing that ranking papers based on the raw number of citations leads to large biases in favor of some PACS codes, while ranking using the relative indicator allows a fair comparison among different fields. We make available, for reference, a table with the average number of citations for each category and each publication year since 1985. We finally present further evidence of the relevance of the rescaling procedure by considering all authors who published in APS journals between 1985 and 2006. There is of course a correlation between the number of raw citations of an author and her/his relative indicator, but large deviations are possible. The problem is even more remarkable when the relative indicator includes a rescaling with the number of authors of each paper, so that the number of citations of a paper are equally split among all co-authors. This highlights that rescaling cites with the number of authors is a crucial issue in citation analysis.

II. DATA SETS

We consider all papers published in journals of the American Physical Society from 1985 to 2009 [20]. We restrict our analysis only to standard research publications (Letters, Rapid Communications, Brief Reports, and Regular Articles) and exclude other type of published material (Editorials, Reviews, Comments, Replies, and Errata) which may show distinct citation patterns. The journals considered in our analysis are Physical Review Letters, Physical Review A, Physical Review B, Physical Review C, Physical Review D, and Physical Review E. APS journals represent the most important publication outlets in physics and cover all subfields of this discipline. They therefore represent an optimal benchmark for the study of citation patterns of publications within physics [21]. The first year of the temporal range considered has been selected because in 1985 the PACS coding started to be systematically used. We consider only papers classified according to the PACS codes, which are the vast majority (>95% between 1985 and 2009) of all papers published in APS journals. PACS numbers are attributed to papers by authors themselves. This guarantees an optimal classification into fields, overcoming the nontrivial problem of attributing, a posteriori, papers to fields [15,22]. PACS codes are composed of three fields $XX.YY.ZZ$, where the first two are numerical (two digits each) and the third is alphanumerical. For our purpose we consider only the first digit of the XX code, which provides a classification into very broad categories (see Table I). Hence, for example, two papers with PACS codes 05.70.Ln and 02.50.Ey both belong to the category 00, while a paper with PACS number 64.60.Ht is part of the category 60. In general, authors assign to a paper two or three PACS numbers. In our analysis we classify papers only according to their principal PACS number. For each paper the number of cites is obtained from the Web of Science database [23], hence including citations from all other non-APS publications included in the WOS database. The data collection was performed on 15 December 2010. In our analysis we consider only papers which have received at least one citation until the above mentioned date.

III. THE RELEVANCE OF THE PROBLEM: MEAN VALUES AND DISTRIBUTIONS

In Fig. 1 we report for each category the number c_0 of citations received on average by each paper as a function of the publication year. It turns out that there are quite large differences in the values of c_0 depending both on the category considered and on the year. In particular, articles published in the same year in a given field can be cited on average up to three times more than in another field (e.g., PACS 40 vs PACS 50 in 1985). The average number of citations received tends to grow as older publications are considered, with rather limited fluctuations, with the notable exception of the very high peak visible in 1988 for PACS 30, due to the extraordinary popularity (more than 30 000 citations so far) of a single article [24]. In general, it is possible to see that the differences between the c_0 of papers published under the same PACS number but in different years can be very large. For PACS 80, for example, $c_0 \approx 80$ for articles published in 1986 but only $c_0 \approx 8$ for publications of 2008.

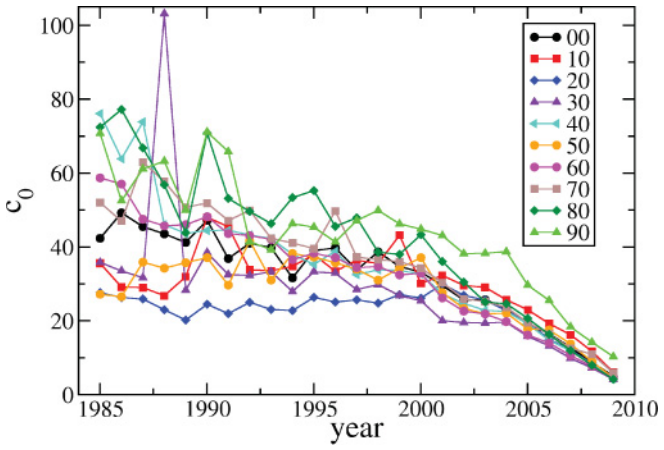

FIG. 1. (Color online) Value of the average number of cites per paper c_0 vs publication year for the different categories.

These data indicate that, although the variations are less pronounced with respect to cross-discipline comparisons [17,18], there are important caveats when trying to compare the number of citations accrued by articles published in APS journals with different PACS codes or publication years.

The same problem is found when one does compare not only the average values but the full distributions. For example, in Fig. 2 we plot the cumulative distribution functions (cdf) of the number of citations for articles in the category 70 for various publication years. It is evident that over the years the distributions change considerably. For instance, more than 200

TABLE I. The 10 categories considered in this study based on the first digit of the first field of the PACS numbers.

PACS	Description
00	General
10	The Physics of Elementary Particles and Fields
20	Nuclear Physics
30	Atomic and Molecular Physics
40	Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics
50	Physics of Gases, Plasmas, and Electric Discharges
60	Condensed Matter: Structural, Mechanical and Thermal Properties
70	Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties
80	Interdisciplinary Physics and Related Areas of Science and Technology
90	Geophysics, Astronomy, and Astrophysics

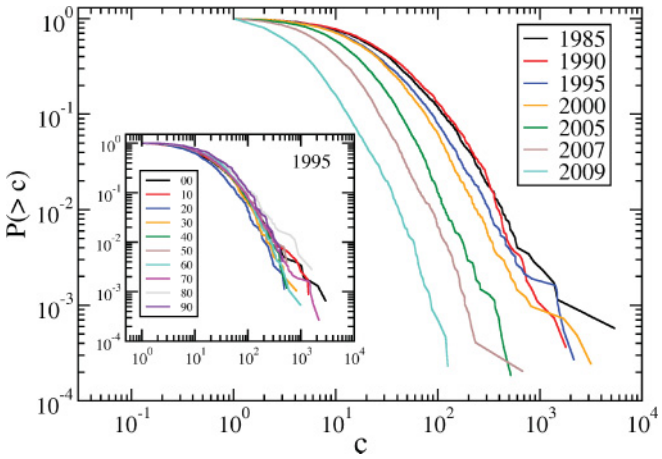

FIG. 2. (Color online) In the main panel we plot the probability $P(>c)$ to observe a paper with more than c citations for category PACS 70 and years, from top to bottom, 1985 (black), 1990 (red), 1995 (blue), 2000 (orange), 2005 (green), 2007 (brown), and 2009 (turquoise). In the inset we plot $P(>c)$ for papers published in the same year 1995 but in different PACS categories: 00 (black), 10 (red), 20 (blue), 30 (orange), 40 (green), 50 (brown), 60 (turquoise), 70 (magenta), 80 (gray), and 90 (violet).

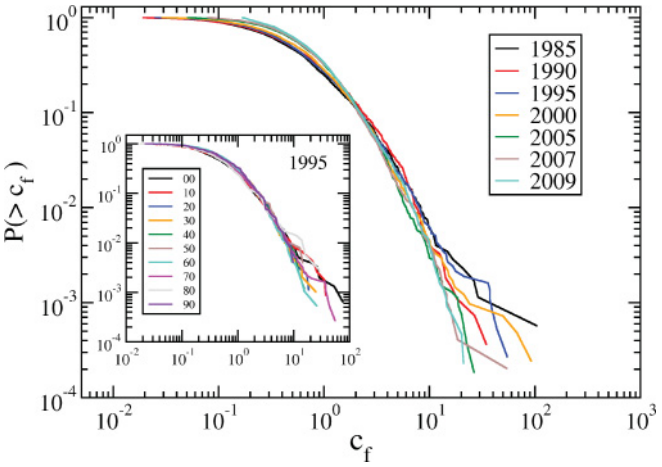

FIG. 3. (Color online) In the main panel we plot the probability $P(>c_f)$ to observe a paper with relative citation count larger than c_f for category PACS 70 and years 1985 (black), 1990 (red), 1995 (blue), 2000 (orange), 2005 (green), 2007 (brown), and 2009 (turquoise). In the inset we plot $P(>c_f)$ for papers published in the same year 1995 but in different PACS categories: 00 (black), 10 (red), 20 (blue), 30 (orange), 40 (green), 50 (brown), 60 (turquoise), 70 (magenta), 80 (gray), and 90 (violet).

citations is not an uncommon result for a paper published in 1985, while it is an extraordinary achievement for one published in 2005. Similarly, citation distributions are also incompatible when one considers the same publication year but different PACS categories (as shown in the inset of Fig. 2 for year 1995).

IV. RELATIVE CITATION RATES AND UNBIASED RANKING

The most natural solution to the problem pointed out in the previous section is the use of relative citation numbers. Let us define the ratio

$$c_f = \frac{c}{c_0}. \tag{1}$$

This quantity measures the success of a paper in terms of citations received compared with other papers in the same category and year. A value $c_f > 1$ (<1) indicates that the paper has been cited more (less) than the average. By definition the average value of c_f is 1, for any category or year, but this is not enough to have a useful unbiased indicator. We need the full distributions to be the same, so that the probability of having any value of c_f is the same, no matter the category or the year. In Fig. 3 we show the distributions of relative citation numbers for the same data of Fig. 2. The comparison between the two figures reveals that the normalization procedure rescales very well all distributions on top of each other, thus allowing a fair comparison among publications in different fields and/or years. Some variation shows up only for recent publication years ($\geqslant$2005) for which citation patterns are generally far from stationary [6]. But also in these cases the difference with respect to the asymptotic shape of the scaled distribution is small. The scaling procedure is able to also remove the bias almost completely for papers published one year ago. A more quantitative assessment of the improvement induced by the use of c_f instead of the raw number of citations c comes from the ranking of articles. We consider papers belonging to several sets and rank them according either to the number of citations c or to the rescaled indicator c_f. We compute then the percentage of publications of each category that appear in the top $z\%$ of

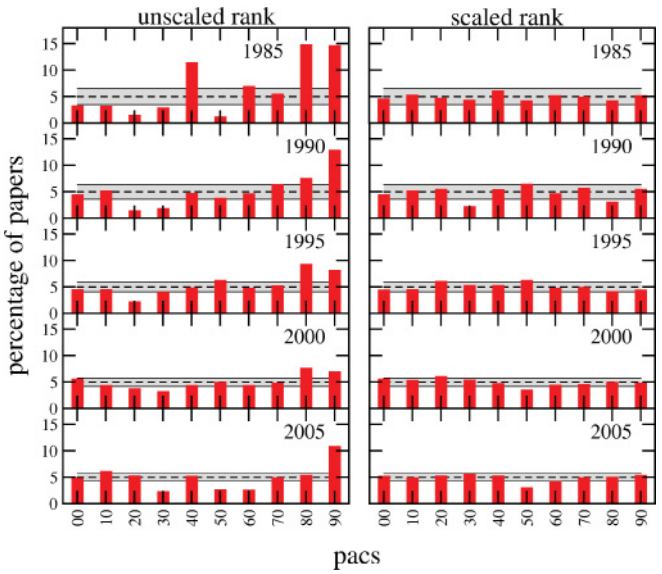

FIG. 4. (Color online) Histograms representing the percentage of papers belonging to the top 5% of the global rank performed according to c (left column) or c_f (right column). In each panel we consider only papers published in a given year and plot the percentage of papers with given PACS category belonging to the top 5% of the global rank. Black dashed lines represent the theoretically expected values (in case of a fair ranking) of the mean, while the gray areas cover the regions corresponding to the mean $\pm$ one standard deviation.

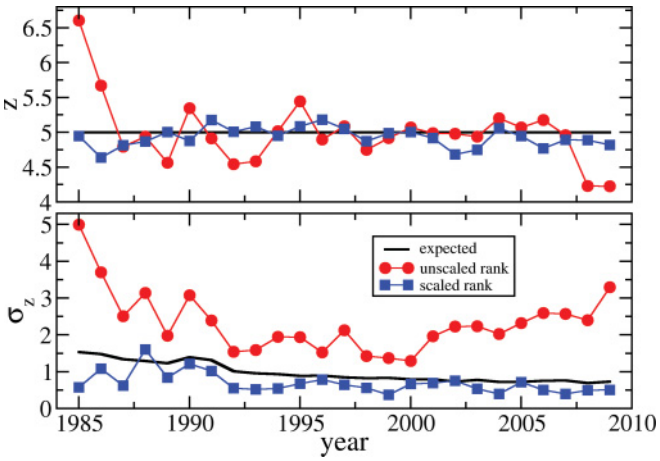

FIG. 5. (Color online) Comparison between the theoretical values (black line) for the mean (top panel) and standard deviation (bottom panel) of bins height in Fig. 4 and the values obtained when the ranking is performed based on c (red circles) or c_f (blue squares).

the global rank. If the ranking is fair the percentage for each category should be around $z\%$ with small fluctuations.

Figure 4 clearly shows that when articles are ranked according to the unnormalized number of citations c there are wide variations among fields. Such variations are dramatically reduced when the relative indicator c_f is used. More quantitatively, assuming that the distributions for the various fields are the same, the expected value of the bin height in

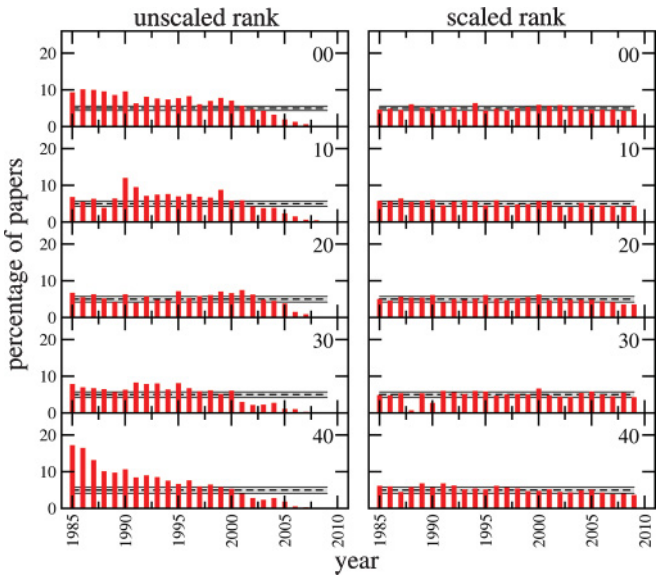

FIG. 6. (Color online) Histograms representing the percentage of papers belonging to the top 5% of the global rank performed according to c (left column) or c_f (right column). In each panel we consider only papers with a given PACS number and plot the percentage of papers published in certain year belonging to the top 5% of the global rank. Black dashed lines represent the theoretically expected values (in case of a fair ranking) of the mean, while the gray areas cover the regions corresponding to the mean ± one standard deviation.

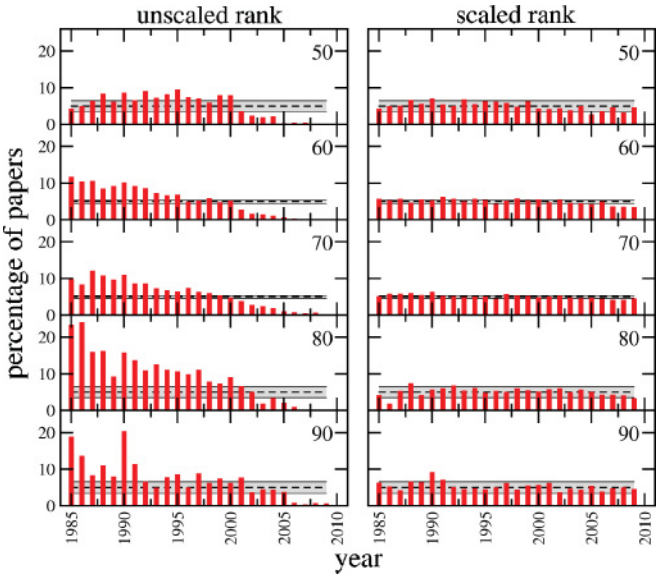

FIG. 7. (Color online) Same as Fig. 6 but for the remaining set of PACS numbers.

Fig. 4 is $z\%$ with a standard deviation

$$\sigma_z = \sqrt{\frac{z(100-z)}{N_c}\sum_{i=1}^{N_c}\frac{1}{N_i}}, \tag{2}$$

where N_c is the number of categories and N_i the number of articles in the ith category. When the ranking is performed according to $c_f = c/c_0$ we find (Fig. 5) a very good agreement with the hypothesis that the ranking is unbiased, while strong evidence that the ranking is biased is found when c is used.

We show in Figs. 6, 7, and 8 the results obtained for different publication years, but fixed PACS categories.

The results presented in Figs. 6–8 demonstrate that the ranking based on c_f is fair, the fluctuations around the expected value $z\%$ being accounted for by finite sample effects. On the contrary, when the raw number c of citations is used, some

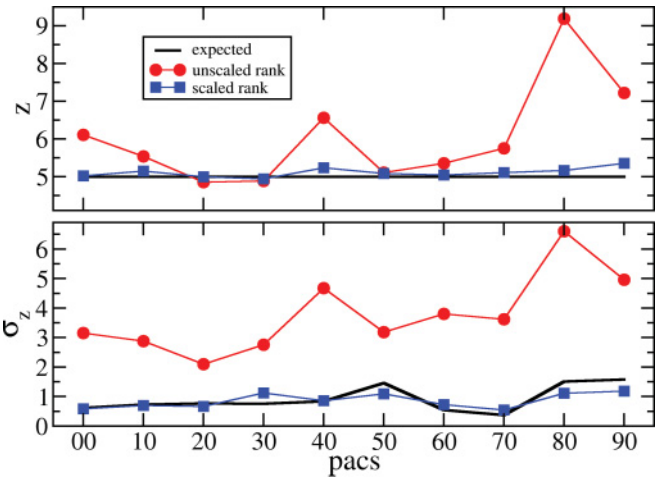

FIG. 8. (Color online) Comparison between the theoretical values (black line) for the mean (top panel) and standard deviation (bottom panel) of bins height in Fig. 6 and the values obtained when the ranking is performed based on c (red circles) or c_f (blue squares).

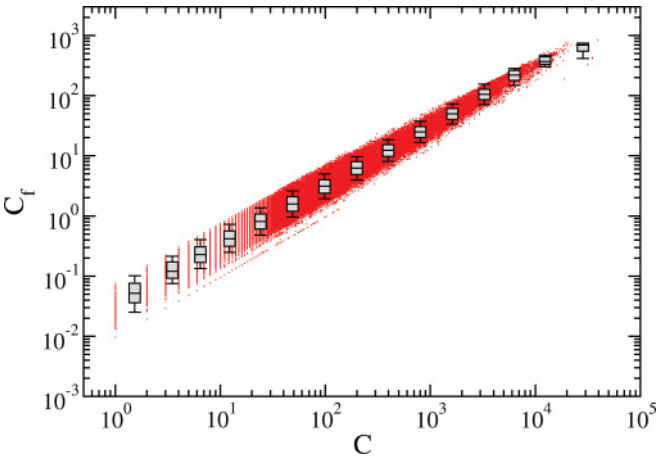

FIG. 9. (Color online) Scatter plot of the values of C and C_f for each author who published in APS journals between 1985 and 2006. For clarity data have been logarithmically binned and for each bin we plot the values of c_f corresponding to the 90% confidence intervals (error bars), 25% confidence intervals, (boxes) and median (horizontal bars inside boxes).

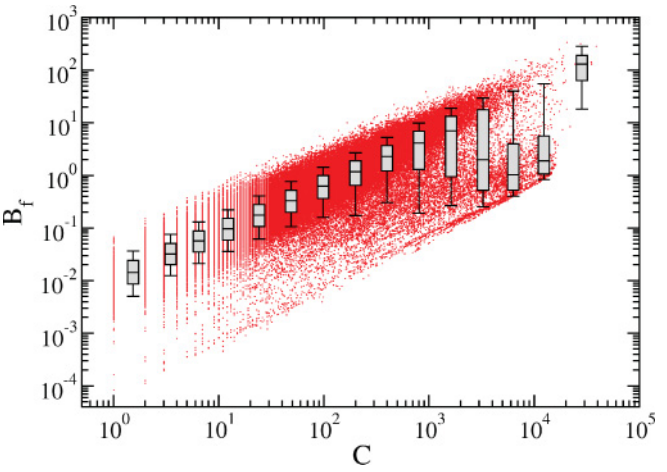

FIG. 10. (Color online) Scatter plot of the values of C and B_f for each author who published in APS journals between 1985 and 2006. For clarity data have been logarithmically binned and for each bin we plot the values of B_f corresponding to the 90% confidence intervals (error bars), 25% confidence intervals (boxes), and median (horizontal bars inside boxes).

sets are largely overrepresented and others underrepresented in the top 5% group. Similar results (not shown) are obtained for other values of z.

V. RENORMALIZED CITATION COUNTS FOR AUTHORS

In this section we present a practical application to the comparison of individual researchers, showing how the rescaled indicator c_f is correlated but far from being equivalent to the raw number of citations c. We have considered all papers published in APS journals from 1985 to 2006 and labeled with PACS numbers and indicated, for the generic paper i, with $c(i)$ and $c_f(i)$ the number of citations received and the relative indicator, respectively. We have then identified for each author a the set $\{a\}$ of their publications[1] and computed the total number of citations $C^{(a)} = \sum_{i\in\{a\}} c(i)$ and the corresponding total value of the relative indicator $C_f^{(a)} = \sum_{i\in\{a\}} c_f(i)$. Figure 9 reports for each author the value of the total rescaled indicator C_f as a function of the total number of raw citations C. The correlation is very good $r \approx 0.98$, but not perfect. For a single value of one indicator there are often values of the other spanning one order of magnitude, indicating again that authors with a very different number of raw citations may be equivalent once differences across subfields and time are considered. An even more striking result is obtained when the relative indicator is further rescaled by the number of authors for each paper: $B_f^{(a)} = \sum_{i\in\{a\}} c_f(i)/N(i)$, where $N(i)$ is the number of authors of the ith paper. This rescaling is aimed at taking into account multiple authorship [26–28] and it is based on the assumption that all authors contribute equally to each publication.[2] In this case the range of variations of B_f for each value of C is increased (the r value is reduced to $r = 0.49$), reaching in many cases two orders of magnitudes. Figure 10 leads to the striking observation that a researcher with 10^4 citations may have a number of rescaled citations equal to a colleague whose publications have been cited only less than 100 times! This result indicates that the normalization of citation counts with respect to the number of paper authors is a critical issue, which can completely overturn the comparison of citation performances of different researchers. The current common habit of attributing all citations to all authors, with no normalization, is unfair and it encourages the misconduct of inflating author lists with individuals who did not actually contribute to the work.

VI. CONCLUSIONS

In this paper we have investigated the possibility of comparing in a fair manner the citations of papers published in APS journals in different years and/or in different fields of physics. We have shown that the raw number of citations is not a suitable indicator, since there are remarkable differences depending on the field and on the year of publication. A fair comparison is obtained instead if the relative number of citations c_f (i.e., the number of citations divided by the average number of cites for the same category and year) is considered. The normalization rescales essentially all distributions on top of each other and this is further confirmed by checking that ranking papers according to c_f does not introduce any bias. For completeness we have performed (but not shown here) the same type of analysis by using as a renormalization factor the median value instead of the average. The median is less

[1]This procedure is subject to many potential errors in the identification of authors. However, this problem is not big [25].

[2]This is for sure largely incorrect in many cases, but it is the most reasonable assumption as long as precise statements about individual contributions are not be published for each article.

sensitive to possible extreme events such as the presence of highly cited papers, but dividing the raw number of cites by the median value leads to less fair comparisons and only for sufficiently old publications (<2000).

In this paper we have considered fields as identified by the first digit of the first number in the PACS code. This classifies all papers in physics into 10 very broad categories. In principle one can further pursue this line of investigation considering more refined levels of categorization. A natural next step would be the consideration of 100 distinct categories, each identified by the whole first field of the PACS code. However, the number of papers published each year in each of these categories is typically very small, and this gives rise to huge fluctuations that do not allow us to extract reliable conclusions.

We believe that these results are very important in view of the increasing trend toward quantitative evaluation of research performance. We strongly encourage researchers dealing with such issues to consider relative citation numbers as the basis of all their evaluations. All indicators for sets of publications (individual authors, groups, research institutions) must be constructed based on the relative citation numbers. This is important also for very large sets (e.g., at the institution level) in order to weigh in a balanced manner the contribution of all fields. For this reason the values of the average number of citations c_0 for each category and each year will be available on the web [29], where they will be updated periodically.

APS journals are an important but clearly partial domain of the whole range of dissemination outlets available for researchers in physics. The extension of the work presented here to also include all other journals where research about physics is published is a much needed step toward a more reliable citation-based research performance evaluation.

Finally, let us stress that the attribution of citations of multiauthored papers to individual contributors is a crucial and much overlooked problem. Different ways of dealing with this problem lead to completely different results (as Figs. 9 and 10 demonstrate). It is of great importance that both researchers and journal editors realize the relevance of this issue while more fair ways of attributing citations are devised.

ACKNOWLEDGMENTS

We acknowledge the American Physical Society for providing the data about Physical Review journals. We thank L. A. N. Amaral and M. Cencini for discussions and feedback on the manuscript.

[1] D. J. de Solla Price, Science **149**, 510 (1965).
[2] L. Egghe, and R. Rousseau, *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*, (Springer, Amsterdam, 1990).
[3] H. Moed, *Citation Analysis in Research Evaluation* (Springer, Dordrecht, The Netherlands, 2005).
[4] J. E. Hirsch, Proc. Natl. Acad. Sci. USA **102**, 16569 (2005).
[5] J. Bar Ilan, J. Informetrics **2**, 1 (2008).
[6] M. J. Stringer, M. Sales-Pardo, and L. A. N. Amaral, PLoS ONE **3**, e1683 (2008).
[7] A. M. Petersen, F. Wang, and H. E. Stanley, Phys. Rev. E **81**, 036114 (2010).
[8] M. J. Stringer, M. Sales-Pardo, and L. A. N. Amaral, J. Am. Soc. Inform. Sci. Technol. **61**, 1377 (2010).
[9] M. H. Macroberts and B. R. Macroberts, Scientometrics **36**, 435 (1996).
[10] R. Adler, J. Ewing, and P. Taylor, Stat. Sci **24**, 1 (2009).
[11] L. Bornmann and H.-D. Daniel, J. Docum. **64**, 45 (2008).
[12] L. Bornmann, Annu. Rev. Inform. Sci. Technol. **45**, 199 (2011).
[13] J. Lane, Nature (London) **464**, 488 (2010).
[14] A. Schubert, and T. Braun, Scientometrics **9**, 281 (1986).
[15] L. Leydesdorff and T. Opthof, J. Informetrics **4**, 423 (2010).
[16] L. Waltman *et al.*, e-print arXiv:1003.2167.
[17] F. Radicchi, S. Fortunato, and C. Castellano, Proc. Natl. Acad. Sci. USA **105**, 17268 (2008).
[18] C. Castellano, and F. RadicchiArch. Immunol. Ther. Ex. **57**, 85 (2009).
[19] [http://publish.aps.org/PACS]
[20] [http://publish.aps.org/datasets-announcement]
[21] S. Redner, Phys. Today **58**, 49 (2005).
[22] K. W. Boyack, R. Klavans, and K. Börner, Scientometrics **64**, 351 (2005).
[23] WOS, www.isiknowledge.com
[24] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[25] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani, Phys. Rev. E **80**, 056103 (2009).
[26] D. J. De Solla Price, Science **212**, 987 (1981).
[27] L. Egghe, R. Rousseau, and G. Van Hooydonk, J. Am. Soc. Inform. Sci. **51**, 145 (2000).
[28] J. W. Hsu and D. W. Huang, Phys. Rev. E **80**, 057101 (2009).
[29] filrad.homelinux.org/resources