Why Sirtes's claims (Sirtes, 2012) do not square with reality

Filippo Radicchi[a], Claudio Castellano[b,c]

[a]Departament d'Enginyeria Quimica, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain
[b]Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy
[c]Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy

---

In a recent Letter (Sirtes, 2012), D. Sirtes presents, in an unnecessarily harsh and aggressive tone, two criticisms to a recent publication of ours (Radicchi & Castellano, 2012) in Journal of Informetrics. In our work, we discussed the problem of identifying quantitative bibliometric indicators for comparing in a fair manner publications belonging to different fields or subfields. We proposed a statistical procedure to test the fairness of indicators and then used it to evaluate two recently proposed indicators (Radicchi, Fortunato & Castellano, 2008; Leydesdorff & Opthof, 2010), when applied to a set of papers, those published in journals of the American Physical Society (APS, `www.aps.org`), belonging to all subfields of physics.

Sirtes presents two arguments against our work. The first is the allegation of "circularity". According to Sirtes "whatever the inequality of the samples of the different fields stems from, whether from genuine field-specific differences or from differences in the quality of the samples, the process of rescaling evens them all out. Now the problem becomes clear: As they use the same sample to calculate their average c_0 as they use to test the fairness of their indicator, their test becomes circular and begs the question."

First, let us notice that this criticism has nothing to do with the fairness test. It has to do with one of the indicators the test is applied to. The criticism is rather obscure. The most straightforward interpretation is that, according to Sirtes, by simply dividing the number of citations by their average, all differences between fields are necessarily canceled. This idea is clearly wrong. Nothing guarantees a priori that simply rescaling by the average many distribution functions is enough to make them equal. In the case of citation distributions for different subfields this scaling property turns out to hold, but this is *highly nontrivial*. We reached this conclusion only after having tested the performance of the rescaled indicator with a statistical analysis. It is surprising that Sirtes is so deeply convinced that our rescaling works, that he considers it to be trivial. There is nothing circular in our method. There is instead a short-circuit in Sirtes's argument.

The second criticism is, fortunately, more clearly formulated. According to Sirtes, APS journals cannot be used as benchmark for our fairness test, because their quality is largely different in the various subfields of physics. The fact that indicators based on raw and fractional citation counts fail our fairness test would be only the consequence of the different quality of APS publications in different physics subfields. The proof of this claim is, according to the author, in Fig. 1 of Sirtes (2012).

This proof does not stand a closer scrutiny. There are several important methodological errors is Sirtes's procedure. Let us first remark that Sirtes did not include in the analysis papers from one APS journal, *Physical Review Letters* (PRL), which publishes many highly cited papers in all physics subfields. Judging the quality of APS journals without considering PRL publications is a serious shortcoming that alone makes the argument of Sirtes inconclusive. Secondly, it is clearly untenable the use of the Impact Factor (IF) (Garfield, 2006) as a measure for the impact of all publications in a journal, despite ample empirical evidence that a huge variability exists in the citations accrued by papers published in the same outlet (Seglen, 1997). But the most inappropriate ingredient of Fig. 1 of Sirtes (2012) is the use of the Journal Citation Reports (JCR) subject-categories and the mapping between them and APS journals. JCR categories are known to be an unreliable categorization of scientific papers. According to Pudovkin & Garfield (2002), in JCR classification "Journals are assigned to categories by subjective, heuristic methods. In many fields these categories are sufficient but in many areas of research these classifications are crude and do not permit the user to quickly learn which journals are most closely related." When using JCR categories for classifying papers, articles are assigned to categories automatically, depending on where they have been published. In doing that, however, a major problem arises because many journals belong to more than one category. For example, papers published in *Physical Review A (PRA)*, automatically belong both to "Physics, atomic, molecular & chemical" and to "Optics". This means that the articles, which make (according to Sirtes) PRA a second rate publishing venue in "Physics, atomic, molecular & chemical" and a top journal in “Optics”, are the same! In our paper, we decided to use *Physics and Astronomy Classification Scheme* (PACS, `publish.aps.org/PACS`) numbers as the categorization for our benchmark exactly to overcome these problems with the JCR categorization. PACS numbers are author-provided and attributed to individual articles, so that they solve the two main problems of JCR categorization. Unfortunately, in the assessment of the relative quality of APS journals with respect to others, the problem with JCR categories cannot be fully eliminated, because, while APS papers can be attributed to JCR categories based on their PACS numbers, non APS publications are classified with the automatic (and unreliable) JCR procedure.

Keeping this caveat in mind, we now proceed to test empirically Sirtes's claim, by measuring the correlation between the prestige of APS papers in a certain subfield and the results of our fairness test. Sirtes claims that such a correlation exists for the indicators based on raw citations and on fractional citations.

As discussed above the mapping between JCR categories and the ten PACS categories considered in Radicchi & Castellano (2012) is a nontrivial problem. We consider the seven JCR subject-categories already analyzed by Sirtes, plus the subject-category “Astronomy & Astrophysics”. We collected from Web of Science (WoS, `isiknowledge.com`) all papers published in year 2000 in journals that are classified in one of these categories, plus all papers published in 2000 in APS journals (including PRL). We consider for consistency only publications that are classified as “Articles” by WoS, and for each of them we retrieve the information about the number of citations accumulated. Data were collected between Feb. 29 and Apr. 4, 2012. To create the mapping between JCR categories and PACS category, we first associate APS papers to JCR subject-categories based on the information given by the first two digits of their principal PACS number. Then, we associate to each JCR category the PACS category where most papers were published in 2000. See Table 1 for the detailed mapping. Note that, while in some cases this mapping is obvious, there are also cases such as “Physics, Condensed matter” in which many PACS with different first digit are placed together. In such a case the attribution to the category 70 is due to the fact that 67% of the papers had PACS code beginning with 7, the 30% of the papers had PACS code beginning with 6, and the remaining papers had PACS code beginning with 4 or 8.

In order to quantify the prestige of APS papers within each JCR subject-category we select the most cited $z\%$ papers appeared in year 2000, irre-

JCR category	PACS first two digits	PACS first digit
PHYSICS, MATHEMATICAL	02 03 04 05 89	00
PHYSICS, PARTICLES & FIELDS	11 12 13 14	10
PHYSICS, NUCLEAR	21 23 24 25 26 27 28 29	20
PHYSICS, ATOMIC, MOLECULAR & CHEMICAL	31 32 33 34 35 36 37 82	30
OPTICS	41 42	40
PHYSICS, FLUIDS & PLASMAS	47 51 52	50
PHYSICS, CONDENSED MATTER	44 61 62 63 64 65 66 67 68 71 72 73 74 75 76 77 78 79 81 83 84 85	70
ASTRONOMY & ASTROPHYSICS	95 96 97 98	90

Table 1: Mapping between JCR categories (first column), first two digits of the of PACS codes (second column), and most frequent first digit of the PACS number (third column). Some first numbers of PACS codes do not appear in the mapping, because they cannot be univocally attributed to a single JCR category.

spectively of their journal of publication, and calculate the proportion $p_{i,z}$ of APS papers that are part of this top cited group. For $z = 10$ for example, we find that the 18% of APS papers are part of the top list in "Astronomy & Astrophysics" while this figure is 25% for "Optics" and 15% for "Physics, Atomic, molecular & chemical".

We then quantify the relative error of each citation indicator, with respect to our fairness test. We take, from Fig. 3 in Radicchi & Castellano (2012), the values $r_{i,z}$ indicating the percentage of papers in PACS category i belonging to the z% globally most cited papers. The relative deviation is given by $q_{i,z}^{(indicator)} = (r_{i,z} - \bar{r}_{i,z})/w_{i,z}$, where $\bar{r}_{i,z}$ is the median value of the proportion predicted by our model and $w_{i,z}$ is the width of the 90% confidence interval (See Fig. 3 in Radicchi & Castellano (2012)) [1].

Using these quantities, we are now able to test the validity of Sirtes's claim. In Fig. 1 we plot $q_{i,z}^{(indicator)}$ vs $p_{i,z}$ for all citation indicators (unscaled, fractional, and rescaled) considered in our previous analysis. The values of the correlation coefficients are reported in Table 2. They are all within the 90% confidence interval expected for a random configuration with same size, which is roughly between -0.6 and 0.6 in all cases. Thus there is no statistically

[1] Notice that the values of z in the determination of $p_{i,z}$ and in $q_{i,z}^{(indicator)}$ are in principle independent, and could also be taken different.

z	R_{unsc}	R_{frac}	R_{resc}
1	0.17	0.54	0.49
2	-0.10	0.20	0.06
5	-0.27	0.31	-0.19
10	-0.16	0.17	0.53
20	-0.01	0.42	-0.33
30	-0.05	0.27	-0.44

Table 2: Values of the correlation coeffients for plots in Fig. 1 for unscaled (R_{unsc}), fractional (R_{frac}) and rescaled (R_{resc}) citation indicators. No value of the correlation coefficient (either positive or negative) is significant at 5% confidence level (the 90% confidence interval is roughly between -0.6 and 0.6 in all cases).

significant evidence of correlation. The failure of the indicators based on raw citations and on fractional citations in (Radicchi & Castellano, 2012) cannot be explained by variations of the prestige of APS papers in the different subfields.

In conclusion, the analysis presented by Sirtes contains serious methodological flaws. Even if such flaws are removed, a statistically grounded approach, presented here, contradicts Sirtes's claims. The present results support instead the validity of the work by Radicchi & Castellano (2012).

Sirtes, D. (2012). Finding the Easter eggs hidden by oneself: Why Radicchi and Castellano's (2012) fairness test for citation indicators is not fair. Journal of Informetrics 6(), 448–450. doi:10.1016/j.joi.2012.01.008

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. Journal of Informetrics, 6(1), 121–130. doi:10.1016/j.joi.2011.09.002

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences, 105(45), 17268–17272. doi:10.1073/pnas.0806977105

Leydesdorff, L., & Opthof, T. (2010). Normalization at the field level: Fractional counting of citations. Journal of Informetrics, 4(4), 644–646. doi:10.1016/j.joi.2010.05.003

Garfield, E. (2006). The history and meaning of the journal impact factor. The Journal of the American Medical Association, 295(1): 90–93. doi:10.1001/jama.295.1.90

Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. British Medical Journal, 314(7079): 498502. doi: 10.1136/bmj.314.7079.497

Pudovkin, A.I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. Journal of the American Society for Information Science and Technology, 53(13): 1113-1119. doi:10.1002/asi.10153

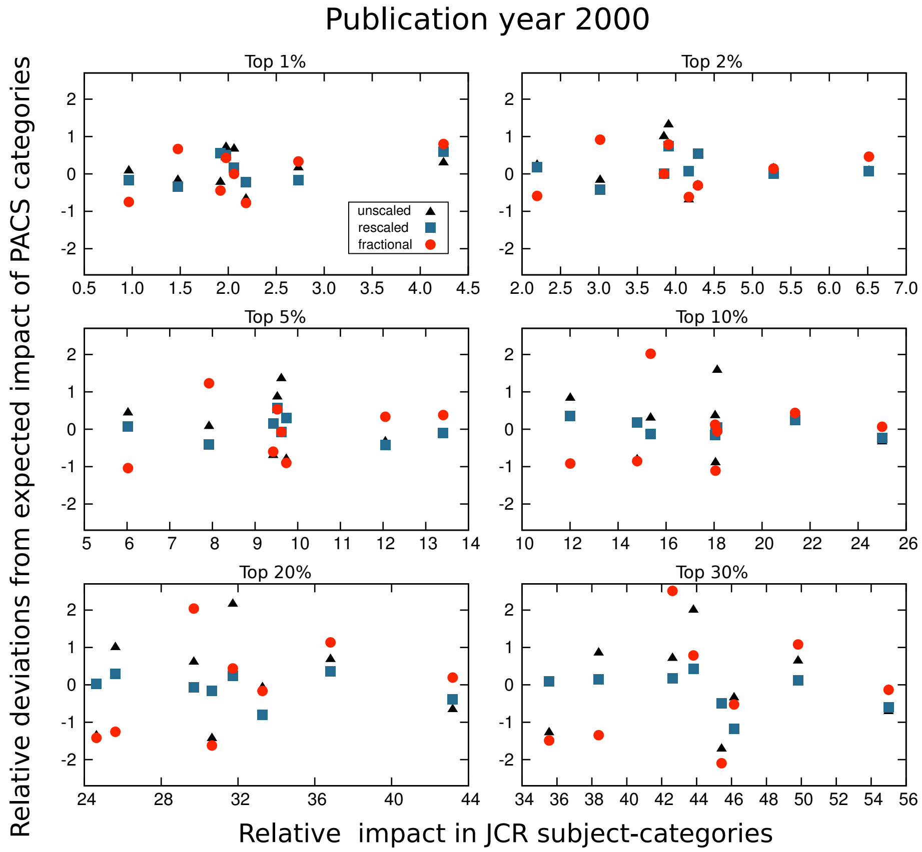

Figure 1: For a given value of z and each PACS category, we quantify the relative deviation from the expected proportion value predicted by Eq.(2) of Radicchi & Castellano (2012). These values are then plotted against the relative impact of APS papers in their JCR subject-category of reference. We consider the same values of z used in our previous analysis (Radicchi & Castellano, 2012).