Physica A 438 (2015) 568–578

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Coherent measures of the impact of co-authors in peer review journals and in proceedings publications

Marcel Ausloos *

Royal Netherlands Academy of Arts and Sciences, Joan Muyskenweg 25, 1096 CJ Amsterdam, The Netherlands
School of Management, University of Leicester, University Road, Leicester LE1 7RH, UK
GRAPES, r. Belle Jardinière, 483, B-4031 Liege, Wallonia-Brussels Federation, Belgium

HIGHLIGHTS

- This paper focuses on the coauthor effect in different types of publications.
- Unexpected relationships are found between coauthor and leading investigator.
- An empirical power law is found between the number of joint publications of an author and the rank of a coauthor.
- Interpretation is based on bibliometrics indices.
- The findings suggest an immediate test of coherence of scientific authorship in scientific policy processes.

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form 5 June 2015
Available online 7 July 2015

Keywords:
Coauthor core value index
Proceedings
Peer review
Coauthorship
Power law relationship

ABSTRACT

This paper focuses on the coauthor effect in different types of publications, usually not equally respected in measuring research impact. *A priori* unexpected relationships are found between the total coauthor core value, m_a, of a leading investigator (LI), and the related values for their publications in either peer review journals (j) or in proceedings (p). A surprisingly linear relationship is found: $m_a^{(j)} + 0.4\, m_a^{(p)} = m_a^{(jp)}$. Furthermore, another relationship is found concerning the measure of the total number of citations, A_a, i.e. the surface of the citation size-rank histogram up to m_a. Another linear relationship exists : $A_a^{(j)} + 1.36\, A_a^{(p)} = A_a^{(jp)}$. These empirical findings coefficients (0.4 and 1.36) are supported by considerations based on an empirical power law found between the number of joint publications of an author and the rank of a coauthor. Moreover, a simple power law relationship is found between m_a and the number (r_M) of coauthors of an LI: $m_a \simeq r_M^{\mu}$; the power law exponent μ depends on the type (j or p) of publications. These simple relations, at this time limited to publications in physics, imply that coauthors are a "more positive measure" of a principal investigator role, in both types of scientific outputs, than the Hirsch index could indicate. Therefore, to scorn upon co-authors in publications, in particular in proceedings, is incorrect. On the contrary, the findings suggest an immediate test of coherence of scientific authorship in scientific policy processes.

1. Introduction

In recent years, studies of complex systems have become widespread among the scientific community, specially in the statistical physics one. Many examples, e.g., Refs. [1–3], pertain to social phenomena in general, indicating that physicists

* Correspondence to: Royal Netherlands Academy of Arts and Sciences, Joan Muyskenweg 25, 1096 CJ Amsterdam, The Netherlands.
E-mail address: marcel.ausloos@ulg.ac.be.

http://dx.doi.org/10.1016/j.physa.2015.06.010
0378-4371/

have gone far from their traditional domain of investigations [4–7]. Moreover, one very modern topics of investigation is the role of measuring (as accurately and objectively as possibly done as in physics) the value of some scientific production [8,9].

In Ref. [10], it was shown that a Zipf-like law

$$J \propto 1/r, \tag{1}$$

exists, between the number (J) of joint publications (NJP) of a scientist, called for short "leading investigator" (LI) with her/his coauthor(s) (CAs); $r = 1, \ldots,$ is an integer allowing some hierarchical ranking of the CAs; $r = 1$ being the most prolific coauthor with the PI. The number of different coauthors (NDCA) is given by the highest possible rank r_M. Several CAs have often the same NJP with the LI.

It was observed that a hyperbolic (scaling) law is more appropriate, i.e.,

$$J = J_0/r^{\alpha}, \tag{2}$$

with $\alpha \neq 1$, usually such that $\alpha \leq 1$, and often decreases with the number of CAs or with the number of joint publications, e.g. when the number of CAs and when J are "not large". J_0 is a fit parameter, i.e. there is no meaning to $r = 0$.

As the h-index [11–13] "defines" the *core of papers of an author* from the relationship between the number of citations n_c and the corresponding rank r of a paper, through a trivial threshold, i.e. if $n_c \geq r_c$, then $r_c \equiv h$, thus one is allowed also to define the *core of coauthors of a scientist* through a threshold [10], called the m_a-index,

$$m_a \equiv r, \quad \text{as long as } r \leq J. \tag{3}$$

This is a specific measure of the core of the most relevant CAs in a research team, centered on the LI. In brief, in the h-index method, one implicitly assumes that the number of "important papers" of an author, those which are the most often quoted, allows to measure the impact of a researcher [14–17]. No need to discuss lengthily the h-index power, variants, or defects. However, such a citation effect is often due to the activity of a research team, centered on the LI [18–21]. In fact, the size and structure of a temporary or long lasting group is surely relevant to the productivity of an author [16]. In contrast, the m_a index as introduced measures the role of coauthors, *rather than citations*, to indicate the most important coworkers of an LI, allowing to measure the LI team core. Technically, one could thus measure the relevant strength of a research group centered on some leader and measure some impact of research collaboration, e.g., on scientific productivity [22]. The invisible college [23,24] of a PI would become visible, easily quantified, whence pointing out to some selection in the community.

Several other measure definitions can be deduced, as in the h-method, i.e. taking into account the whole surface of the histogram, i.e. the cumulated number of joint publications (NJP)

$$\Sigma \equiv \sum_{r=1}^{r_M} J_r, \tag{4}$$

for the CA with rank r has published J_r publications with the LI. An often discussed part of the histogram is that up to the threshold; it corresponds to the cumulated NJP limited to the core, i.e.

$$A_a \equiv \sum_{r=1}^{m_a} J_r. \tag{5}$$

The notation is reminiscent of the A-index [25–27], in the Hirsch scientific output measurement method of an author. Of course, $A_a / \sum$ gives the relative weight of the core CAs in the cumulated NJP.

Moreover, one can define an a_a-index [10] which measures the surface below the empirical data of the number of joint publications *till the CA of rank* m_a, normalized to m_a, i.e.

$$a_a = \frac{1}{m_a} \sum_{r=1}^{m_a} J_r \equiv \frac{A_a}{m_a}, \tag{6}$$

and similarly the index

$$a_M = \frac{1}{m_a} \sum_{r=1}^{r_M} J_r \equiv \frac{\Sigma}{m_a} \tag{7}$$

measured from the *whole* histogram surface. Obviously, $A_a / \sum \equiv a_a / a_M$. The notations are similar to those of the h-index scheme, where they somewhat measure the average number of citations of papers *in the Hirsch core* [13].

Note that the true mean μ of the J *vs.* r distributions, i.e. the average NJP per CA, is obtained from

$$\mu = \frac{\Sigma}{(NDCA)} \equiv \frac{\Sigma}{r_M}. \tag{8}$$

In practical terms, these indirect measures are attempts to improve the sensitivity of the threshold forced index in order to take into account the number of co-authors whatever the number of joint publications among the most frequent coauthors, and introduce a contrast between the most frequent CAs and the less frequent ones. Indeed, JP has often a mix of different CAs[1] [28]. It has also been observed in previous fits, through Eq. (2), that unusual (long or short) lists of coauthors, as well as the hapax-like CA, i.e. accidental or rare CAs, but with necessarily large r values, have much influence on J_0 and α, and the resulting R^2.

Moreover, it is somewhat commonly accepted that proceedings papers, e.g. resulting from conference presentations, have to be distinguished from peer review journal publications. Miskiewicz [29] has discussed whether such different types of publications have some impact on the core number and on the ranking of CAs. For completeness, note that a complementary question was also examined, i.e. whether a "binary scientific star" – like system implies some deviation from Eq. (2) – the "binary scientific star" (BSS) being defined as the couple formed by an LI and one of his most frequent CAs [30].

In the following sections, an amazingly simple relationship is reported to be found between $m_a^{(jp)}$ and its related value for publications in peer review journals (j) and in proceedings (p), i.e. $m_a^{(j)} + 0.4\ m_a^{(p)} = m_a^{(jp)}$. Moreover, another relationship is found concerning the A_a index, i.e. the surface of the J $vs.$ r histogram up to m_a, i.e., $A_a^{(j)} + 1.36\ A_a^{(p)} = A_a^{(jp)}$. A discussion of other empirical (linear) relations is presented. The illustrative data of the coauthorship features is quickly recalled for the few published cases, in Section 2. In Section 3, some hint is presented on some origin of the, surprising (or unexpected), relationship, and for the coefficient values. The case of anomalous data points is also discussed. Some justification is based on the empirical power laws $J^{(t)} \propto r^{-\alpha^{(t)}}$, [10], emphasizing that $\alpha^{(t)}$ depends on the type (t) of publication, even if only slightly. A bonus (?) is found to be the simple power law relationship between the m_a core value and the number of different coauthors, see an Appendix.

Note that there is at first no reason to predict that a simple relationship will be found between the various quantities here above introduced. In fact an examination of the distributions led to ambiguous results [30]. There is apparently no other previous investigation of this matter. *In fine*, modeling likely requests much more thinking. Section 4 serves as a conclusion on the respective relevance of different types of publications in "evaluating" an LI and his/her CAs, and with some suggestion for future work. Nevertheless, the findings could imply practical considerations on subsequent measurements of publication activities during a career, -as self-citations might do [31].

Note also that several other so called *laws* have been predicted or discovered about relations between number of authors, number of publications, number of citations, fundings, dissertation production, citations, or the number of journals or scientific books, time intervals, etc. [32].

2. Data sample

For the following study and discussion the same LIs as those investigated in previous publications [10,28–30] are considered. The LIs span a large range of scientific research topics, though in statistical physics mainly. They are mentioned by their initials. Most of them are males, except two. They come from mainly Poland, 7, i.e. RW, JMK, AP, KSW, JM, and MM; 4 are from the "western world", HES, DS, MA, and PC. They are half several senior (JMK, AP, HES, DS, MA, PC) and half rather junior scientists. In previous reports [10,28–30], their publication list has been summarized and is thus not recalled here. Beside the LIs, 4 BSS cases [30], i.e. so called HSSH, HSSB, MARC and MANV have been made for further completing the up to-day rather rare data. This leads to examine 15 cases. *A priori*, the data does not seem to be specifically biased.

The best power law fits, through Eq. (2), α and m_a value, and the distribution main statistical characteristics, i.e., the mean μ and r_M, are given in Table 1. This well illustrates the similarity in behavior, but points to differences to be examined next in more detail. One may expect, from a general point of view, that the subsets, i.e. joint publications in peer review journals (j) and in proceedings (p), might have some influence on characteristics of those for the whole (jp) set since they form the structure. However, the problem is highly nonlinear in essence, since the "rank" is not a usual variable. Nevertheless, in line with modern statistical analysis, and in order to detect some substructure, several fits can be attempted, i.e. power law, exponential, logistic,..., and polynomial, the most simple being the linear one.

The deduced values of $\sum$, A_a, a_M and a_a are given for the three sets, i.e. (jp), (j) and (p), in Table 2. The empirically found laws are presented in Figs. 1–4, for the m_a, A_a, a_M, a_a quantities.

A simple relation is found between the core measures, i.e.

$$m_a^{(j)} + 0.414\ m_a^{(p)} = m_a^{(jp)} \tag{9}$$

with a high regression coefficient R^2, i.e. $\sim$0.894, and between the histogram surfaces below the corresponding core measures

$$A_a^{(j)} + 1.36\ A_a^{(p)} = A_a^{(jp)} \tag{10}$$

with a very high regression coefficient R^2, i.e. $\sim$0.998, as seen in Figs. 1–2 respectively. In both cases, a classical two parameter linear fit has been attempted. It has been found that $m_a^{(jp)} = m_a^{(j)} + 0.452\ m_a^{(p)} - 0.321$ and $A_a^{(jp)} = A_a^{(j)}$

[1] The order of authors is at this level not discussed.

Table 1
Summary of direct data values for 11 LIs and 4 BSSs: m_a is the core measure [10]; α is the exponent of the empirical power law, Eq. (2); μ is the mean of the distribution (J *vs*. r); r_M the total number of different CAs (NDCA); always distinguishing among of joint publications, the total (jp) sum, the journals (j) and the "proceedings" (p).

	$m_a^{(jp)}$	$m_a^{(j)}$	$m_a^{(p)}$	$\alpha^{(jp)}$	$\alpha^{(j)}$	$\alpha^{(p)}$	$\mu^{(jp)}$	$\mu^{(j)}$	$\mu^{(p)}$	$r_M^{(jp)}$	$r_M^{(j)}$	$r_M^{(p)}$
HES	26	20	15	1.135	0.999	1.045	6.569	4.67	5.136	592	568	242
DS	12	12	3	0.796	0.535	0.688	2.725	2.578	1.565	280	268	46
MA	20	15	10	1.102	1.029	0.86	4.872	3.865	3.041	319	273	168
PC	4	3	3	0.87	0.94	0.67	3.097	2.684	2.0	302	129	24
RW	6	4	4	0.743	0.767	0.561	2.75	1.94	2.71	46	34	23
JMK	5	4	3	0.787	0.702	0.618	2.707	1.714	2.04	41	35	25
AP	6	5	2	0.94	0.64	0.89	2.872	2.622	1.5455	47	45	11
DG	2	2	2	0.547	0.755	0.239	1.13	1.75	1.05	104	7	99
KSW	3	3	1	0.715	1.255	0.594	2.13	2.0	1.67	21	21	3
JM	2	2	1	0.63	0.67	0.67	1.75	1.71	1.33	14	12	1
MM	3	2	2	0.536	0.428	0.521	1.515	1.3125	1.45	33	16	20
HSSH	16	11	10	1.074	0.934	0.974	5.602	3.87	3.895	196	169	114
HSSB	15	10	10	1.064	0.922	0.969	5.104	3.549	3.766	214	176	114
MARC	11	9	7	0.985	0.893	0.887	3.81	3.07	2.958	147	114	71
MANV	5	3	3	0.835	0.755	0.67	2.60	2.143	1.833	40	28	24

Table 2
Summary of indirect data values for 11 LIs and 4 BSSs: $a_M \equiv \Sigma/m_a$, where m_a is the CA core measure; Σ is the surface below the histogram (J *vs*. r), i.e. TNCA; A_a is the surface below the J *vs*. r histogram, limited to the core value m_a; $a_A \equiv A_a/m_a$; each value for the total (jp) number of joint publications, journals (j) or "proceedings" (p).

	$a_M^{(jp)}$	$a_M^{(j)}$	$a_M^{(p)}$	$\Sigma^{(jp)}$	$\Sigma^{(j)}$	$\Sigma^{(p)}$	$A_a^{(jp)}$	$A_a^{(j)}$	$A_a^{(p)}$	$a_a^{(jp)}$	$a_a^{(j)}$	$a_a^{(p)}$
HES	149.6	131.95	83.3	3889	2639	1250	1625	895	549	62.5	44.75	36.6
DS	63.58	57.58	24	763	691	72	259	229	16	21.58	19.08	5.33
MA	77.85	70.33	50.2	1557	1055	502	810	482	221	40.5	32.13	22.1
PC	24.75	17	16	99	51	48	46	29	16	11.5	9.67	5.33
RW	21.5	16	16.25	129	64	65	64	21	33	10.67	10.2	8.25
JMK	22.2	15	17	111	60	51	39	21	16	7.8	5.25	5.33
AP	22.5	23.6	8.5	135	118	17	64	51	7	10.67	10.2	3.5
DG	59	7	52	118	14	104	14	7	7	7	3.5	3.5
KSW	16.33	14.67	5	49	44	5	18	15	3	6	5	3
JM	13.5	11.5	2	27	23	2	12	10	3	6	5	3
MM	16.67	10.5	14.5	50	21	29	16	7	8	5.33	3.5	4
HSSH	68.625	59.45	44.4	1098	654	444	524	236	204	32.75	21.45	20.4
HSSB	72.53	62.1	46.7	1088	621	467	469	202	191	31.27	20.2	19.1
MARC	50.91	38.89	30	560	350	210	280	151	97	25.45	16.778	13.86
MANV	20.8	20	14.667	104	60	44	52	26	19	10.4	8.667	6.33

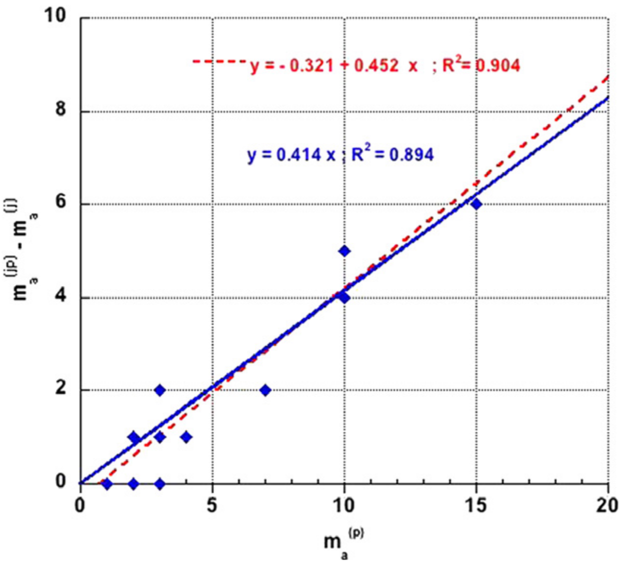

Fig. 1. Empirical proof of the linear relationship between the total CA core value $m_a^{(jp)}$ and the corresponding ones, but distinguishing between peer review journals (j) and "proceedings" (p) papers, i.e. $m_a^{(j)} + 0.414\ m_a^{(p)} \simeq m_a^{(jp)}$; $R^2 = 0.894$; the dashed line indicates the best possible two parameter linear fit; several data points overlap each other.

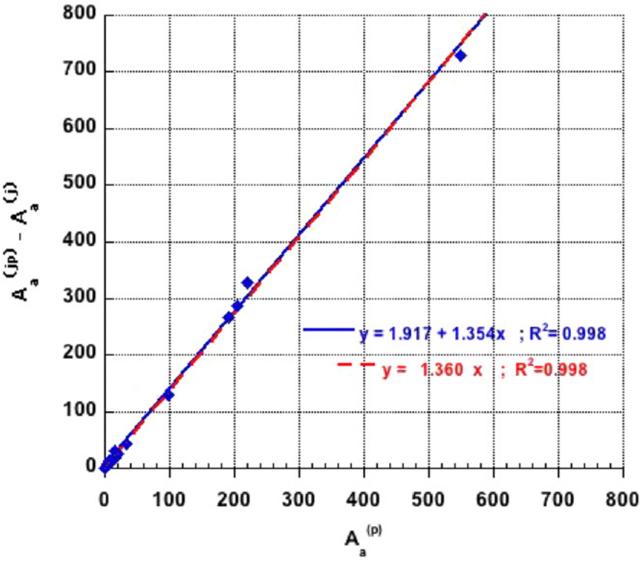

Fig. 2. Empirical proof of the linear relationship, $A_a^{(jp)} \simeq A_a^{(j)} + 1.36\ A_a^{(p)}$, between A_a's, i.e. the (Number of joint publications–Coauthor rank) histogram *J vs. r* surface below m_a [10], and the A_a corresponding values for peer review journals (j) and "proceedings" (p); $R^2 = 0.998$; the dashed line indicates the best possible two parameter linear fit; several data points overlap each other.

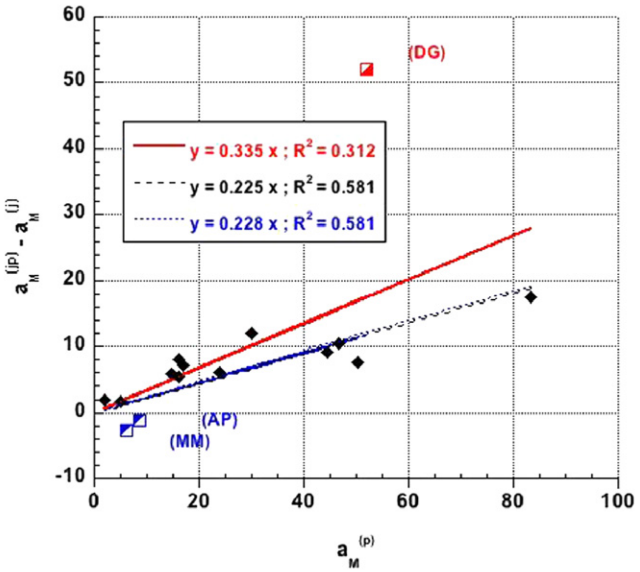

Fig. 3. Empirical linear relationship, $a_M^{(jp)} \simeq a_M^{(j)} + 0.225\ a_M^{(p)}$, between the whole a_M's and the corresponding ones for joint papers with coauthors in peer review journals (j) and "proceedings" (p); $R^2 \simeq 0.581$; if either the (DG) and the (AP) and (MM) points are not considered in the fit, as being possible "outliers", the relationship numerical values are slightly modified. N.B. The best possible linear two parameter fits are not shown though they have a higher R^2, see text for values, but the abscissa at the origin can hardly be interpreted.

$+\ 1.354\ A_a^{(p)} + 1.917$ respectively, with the corresponding R^2 values being equal to 0.904 and 0.998. Such fits are shown in Figs. 1–2.

Let it be observed that one has necessarily

$$\Sigma^{(j)} + \Sigma^{(p)} = \Sigma^{(jp)} \tag{11}$$

which is nothing else that a normalization condition on the NJP, as exemplified in Table 2. Subsequently, a_M has been examined. The result is reported in Fig. 3. One finds

$$a_M^{(j)} + 0.225\ a_M^{(p)} = a_M^{(jp)}, \tag{12}$$

but with a low $R^2 \simeq 0.581$. More discussion, stemming from the presence of anomalous data, so called "outliers", is found in Section 3.

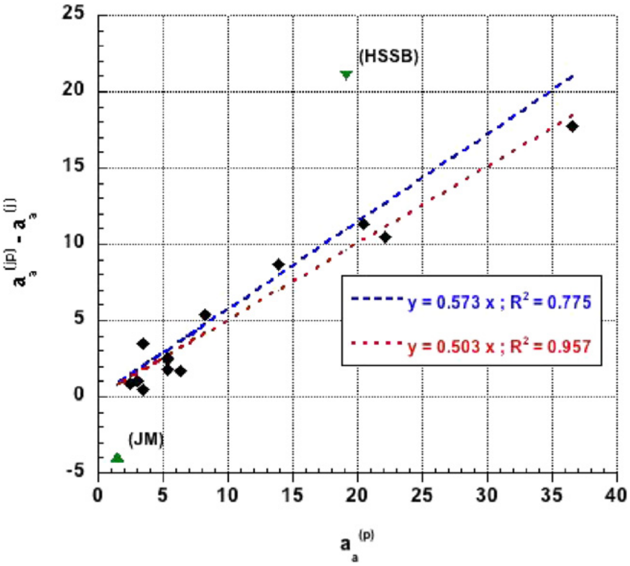

Fig. 4. Empirical proof of the linear relationship, $a_a^{(jp)} \simeq a_a^{(j)} + 0.503\ a_a^{(p)}$, between a_a values resulting from the (Number of joint publications (jp)–Coauthor rank) histogram, J *vs.* r, surface below the core value m_a, and that of peer review journals (j) and "proceedings" (p); $R^2 = 0.957$, when (JM) and (HSSB) points are not considered, as being outliers, see text; inclusion of such points are shown by the blue dashed line; the best linear two parameter fits are not shown because the abscissa at the origin can hardly be interpreted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, a fine linear fit is found for a_a as seen in Fig. 4

$$a_a^{(j)} + 0.503\ a_a{}^{(p)} = a_a{}^{(jp)}, \tag{13}$$

with $R^2 \simeq 0.957$, even with the presence of anomalous (outlier) data, as discussed in Section 3.

3. Discussion

3.1. Outliers

First let us discuss the cases of a_M and a_a, before introducing an estimate of the numerical coefficients. Visual inspection shows that there are a few data points looking like outliers[2] in Figs. 3 and 4. They are called (DG), (AP), and (MM) on one hand and (JM) and (HSSB) on the other hand.

For the a_M case, Fig. 3, it is remarked that (DG) has a specially high number of CAs, i.e. 93, having only one ("proceedings") joint publication with DG. This stems from some work by DG in high-energy physics before he turned towards statistical mechanics research. If such a data point is included in a proportionality fit between $a_M^{(jp)} - a_M^{(j)}$ and $a_M^{(p)}$, one obtains a proportionality coefficient equal to 0.335, and a R^2 value equal to 0.312. Concerning (AP) and (MM), it is observed that $a_M^{(jp)} - a_M^{(j)} \le 0$, which may at first appear awkward and unattractive. Note that the respective values of $\sum$ and m_a seem "reasonable". However, the origin of the negative value is likely attributable to the fact that AP and MM have very few "proceedings" papers, having mainly concentrated their research output into peer review journals, and few papers resulting from scientific meetings. In some sense, through these two authors, one point out to the effect of duplicate-like papers.

The comment on the "(DG) effect" implies that one should deduce that the proportionality coefficient is likely to be dependent on the research field. On the other hand, the comment on the "(AP)–(MM) effect" indicates that one should allow for a negative value of $a_M^{(jp)} - a_M^{(j)}$, and propose that the coefficient to be accepted is 0.228 rather than 0.225.

For completeness, a classical two parameter linear fit has been made either taking into account all data points, or removing outliers. The following relations have been obtained, for the a_M

- taking into account all data points, $y = 0.255 + 0.33x$, with $R^2 = 0.349$,
- without the (DG), (AP) and (MM) points, $y = 2.964 + 0.164x$, with $R^2 = 0.769$,
- without the (DG) point, $y = 1.232 + 0.197x$, with $R^2 = 0.718$.

2 Interesting comments on outliers can be found in Refs. [33,34].

For the a_a cases, see Fig. 4, HSSB is seen to have a very high $a_a^{(jp)} - a_a^{(j)}$ positive value, while for JM, one obtains $a_M^{(jp)} - a_M^{(j)} \leq 0$. It is fair to emphasize that there is a similarity between a_a and a_M cases. However, the deductions originate from different surfaces, i.e. below the cores in the a_a cases, to be contrasted with the whole surfaces for the a_M cases.

Since HSSB has almost equal NJP, with their main CAs, in either p or j set, one might wonder if these are duplicate results, since they imply the same and main CAs. The case of JM shows that this is not a duplicate finding: indeed, on the contrary, JM has very few p-type publications with his/her main CAs. Moreover, the different "behavior" of HSSB and JM enlightens the fact that JM has rather a concentration of scientific output in peer review journals with his/her main CAs (like for AP and MM, in fact).

For completeness, a classical two parameter linear fit has been made taking into account all data points, or removing outliers. The following relations have been obtained, for the a_M:

- taking into account all data points, $y = -0.671 + 0.608x$, with $R^2 = 0.775$,
- without the (HSSB) and (JM) points, $y = -0.07 + 0.507x$, with $R^2 = 0.957$.

Note that one should not be impressed by the respective R^2 values – obviously depending on the number of data points, and the fact that these are two parameter fits – in contrast to the values given here above and in either Eq. (12) or Eq. (13).

The positive or negative value of the abscissa at the origin can be attributed to the fact that the resulting combination (jp) between the (j) and (p) set is *a priori* highly non-linear. Indeed the various ranks do not sum up, since a CA in one set may appear at two rank values totally unrelated to the resulting rank for the (jp) set.

Finally, in all cases, one may deduce that these (outlier-like) results arise from different scientific (or other) behavior of the respective scientists, but this discussion is outside the realm of the present paper.

3.2. Theoretical estimates

The numerical proportionality coefficients, as well as those resulting from a two linear parameter fit, can be discussed, going to the continuum limit for the respective histograms. Indeed, within a continuum approximation, one has

$$\sum_{r=1}^{r} J_r \to \int_1^r J(r)\,\mathrm{d}r \equiv \int_1^r \frac{1}{r^\alpha}\,\mathrm{d}r = J_0\,[r^{(1-\alpha)} - 1]. \quad (14)$$

Consequently,

$$\sum_{r=1}^{r_M} J_r \equiv \Sigma \to = J_0\,[r_M^{(1-\alpha)} - 1], \quad (15)$$

$$A_a \equiv \sum_{r=1}^{m_a} J_r \to = J_0\,[m_a^{(1-\alpha)} - 1], \quad (16)$$

$$\frac{\Sigma}{m_a} \to = \frac{J_0}{m_a}\,[r_M^{(1-\alpha)} - 1], \quad (17)$$

$$\frac{A_a}{m_a} \to = J_0\,[m_a^{-\alpha} - m_a^{-1}]. \quad (18)$$

Recall that these quantities depend both on the type of publication set and on the LI. For example, for one specific LI, and some publication set, one has from Eq. (16),

$$(A_a/J_0) + 1 = m_a^{1-\alpha}, \quad (19)$$

while from Eq. (15), one finds

$$[\Sigma/J_0\,] + 1 = r_M^{1-\alpha}. \quad (20)$$

Observe that

$$\frac{\Sigma + J_0}{A_a + J_0} = \left(\frac{r_M}{m_a}\right)^{(1-\alpha)} \quad (21)$$

for each LI and each type of scientific publication. Some elementary, but very tedious algebra, can follow. From Eq. (21), one can extract Σ, and rewrite explicitly Eq. (11), in order to obtain a linear relationship between the three A_a quantities, such that one can write

$$A_a^{(jp)} = \Lambda_j\,A_a^{(j)} + \Lambda_p\,A_a^{(p)} + [\,\ldots\,], \quad (22)$$

where each Λ and [...] are functions of r_M and m_a. Moreover, as shown in the Appendix, one has a (surprisingly simple) power law relationship between m_a and r_M, i.e. $m_a = \upsilon\ r_M^{\beta}$, (or $r_M = u\ m_A^{\gamma}$), see Figs. 5–6. Therefore, one can evaluate the

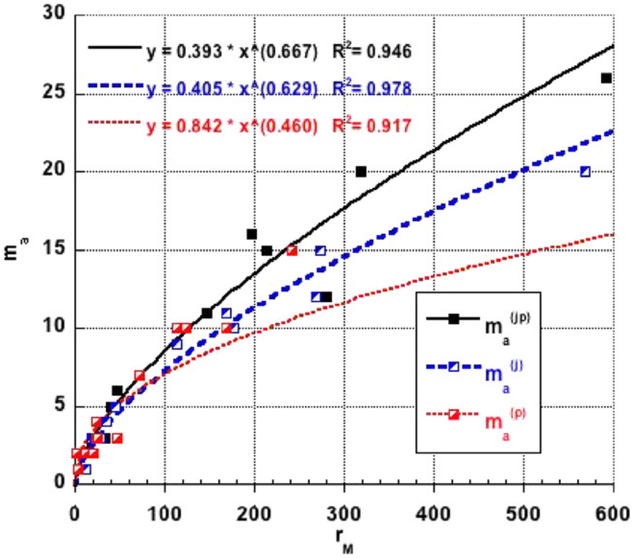

Fig. 5. Empirical proof of the power law relationship, $m_a \simeq r_M^{\beta}$, resulting from the (Number of joint publications (*jp*)–Coauthor rank) histogram, J *vs*. r, surface and that for peer review journals (j) and "proceedings" (p), when the (DG) point is not considered, as being an outlier.

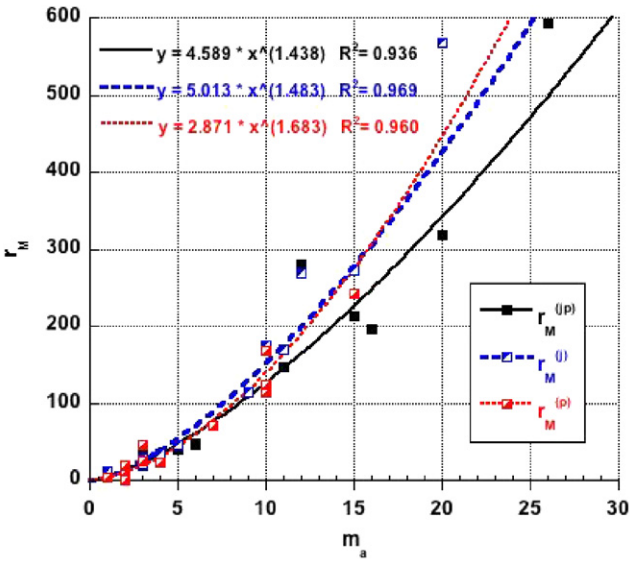

Fig. 6. Empirical proof of the power law relationship, $r_M \simeq m_a^{\gamma}$, when the outlier (DG) point is not considered, for the cases of peer review journals (j) and "proceedings" (p), and the total (*jp*) number of joint publications.

ratio Λ_p/Λ_j for the various cases; it is about $v_p/v_j \sim (2.9/5.0)^{-(2/3)} \sim 1.44$, not too far from 1.36. The complicated [. . .] term can be roughly estimated, according to the various numerical values. It is found rather small with respect to the other terms, whence corroborating the finding in Eq. (10).

To "verify" Eq. (9) is more subtle and complicated. Indeed, one has to start from Eq. (11), in which one substitutes each Eq. (15). This leads to a highly nonlinear relationship of the type

$$M_a^{(jp)} = \Omega_j\, M_a^{(j)} + \Omega_p\, M_a^{(p)} + [\,\dots\,], \tag{23}$$

where

$$M_a \equiv m_a^{\gamma(1-\alpha)}. \tag{24}$$

To extract a linear relationship, analytically, like $m_a^{(jp)} = \omega_j\, m_a^{(j)} + \omega_p\, m_a^{(p)} + [\,\dots\,]$, similar to Eq. (9) seems feasible only if $\gamma(1-\alpha) \equiv (1/k)$, where k is an integer. For the sake of argument, taking $\gamma = 3/2$ and $\alpha = 1/3$, thus $k = 1$, one can estimate according to the values in Tables 1–2 and those in the Appendix that the ratio $\omega_p/\omega_j \sim J_0^{(p)}/J_0^{(j)} \simeq 0.5$, not too far from 0.4 in Eq. (9).

4. Conclusions

In summary, recall that the core m_a of coauthors (CAs) of a leading investigator (LI) is well defined through a criterion similar to the h-index. Through a J $vs.$ r histogram, one describes a CA "impact" according to the number of his/her joint publications J with an LI. It is usually considered that scientific publications in proceedings differ, in various ways, "values", from those in peer review journals. This belief was questioned here above. In fact, one can distinguish the core of coauthors of an LI, according to the type of joint publications. Next, it was wondered whether the relative hierarchy in estimating the value of publications in journals or proceedings can be carried to the core of co-authors. Finally, the question was raised: "is there any numerical proof of the usual belief in a qualitative difference?"

Visually, it appears that the hierarchy exists, i.e. $m_a^{jp} > m_a^j > m_a^p$, but, somewhat surprisingly, the relationship turns out to have a simple analytic form: $m_a^{(j)} + 0.4\ m_a^{(p)} = m_a^{(jp)}$. Moreover, another relationship is found concerning the A_a index, i.e. the surface of the J $vs.$ r histogram up to m_a, i.e., $A_a^{(j)} + 1.36\ A_a^{(p)} = A_a^{(jp)}$. These linear relationships also hold for subsequently derived measures.

The findings have been illustrated gathering data for a dozen or so LIs, and for 4 couples, i.e. publications in which an LI is systematically with a specific CA. Even though the data, about scientists working in the research field of statistical physics, is of finite size, the result does not seem to be biased. A χ^2 test indicates the reliability of the found features. Yet, in one case, an LI, having previously worked in high energy physics and having many (93) CAs for one publication, of the (p)-type, some anomalous behavior occurs. This outlying feature should not distract from the main findings.

The main text analysis points out to the interest of the measures of CA cores according to types of publications. The results of course suggest to investigate other scientific domains. In fact, it can be done through a bonus, the discovery that m_a is tied with the maximum number of CAs of an LI, i.e. r_M, through a simple power law. In so doing, one may observe that outliers are easily found [33,34]—thus removing them leads to a large increase in the χ^2 coefficient, *in fine* giving much weight to the interest of such numerical findings. It is also obvious that the analysis is rather simple for policy makers—when the list of publications is available.

Theoretical work, whence explanations, have been shown not to be trivial, because the systems are highly non-linear. A CA might appear in one type (j or p) of publications, but not in the other type. However, as it is more frequently seen, a CA appears in both types of publication. However, it might be at a quite different ranking. This is the case when a CA does not have "many" publications with an LI. In fact, the ranking of some CA in some type of publication does not simply add up with the ranking in another type of publications. A CA ranking is usually not conserved, but depends on the publication type. Whence, there is *a priori* no reason why a linear relationship should be found between such measures.

However, there is one "normalizing condition", in general not appreciated, which imposes further consideration of such co-authorship measures: the surface of the histogram J $vs.$ r for the whole set of publications is necessarily the sum of the surfaces for the two types of publications. Thus, since the continuum limit, $J \propto 1/r^{\alpha}$, one can measure such surfaces as a function of α and r_M, and later derive an estimate of the numerical coefficients obtained through empirical fits.

Finally, one should not simply consider that these are numerical games. They lead to remarkable proportionality measures of CA roles. It appears that the number of CAs "of interest" for measuring the core of CA of an LI is mainly arising from the joint publications in peer review journals. Indeed, only about ($0.4/1.4 \simeq$) 30% stem from "proceedings". Similarly, it appears that the "contribution" to the number of joint publications by the main CAs is about 50% of the whole.

This implies to elaborate practical considerations on subsequent measurements of publication activities during a career, on the role and effects of coauthors [35–37]. The case of outliers is also of interest, since it carries some weight in estimating α and subsequent numerical coefficients.

These simple relations imply that coauthors are a "more positive measure" of a principal investigator role, in both types of scientific outputs, than the Hirsch index which barely counts the number of citations independently of the co-author (number nor *a fortiori* rank). Therefore, to scorn upon co-authors in publications, in particular in proceedings, is highly incorrect. On the contrary, the findings suggest an immediate test of coherence of scientific authorship in scientific policy processes. This could imply many practical considerations on the role of CAs with respect to an LI and on the respective roles of different types of publications in "measuring" an LI team work. A discussion of criteria based on the above for estimating, e.g. the financing of an LI or a team, is outside the realm of the present paper. Nevertheless, with softwares actually available to policy makers, the development and application of such findings should be easily possible.

Note three final points: (i) each rank-frequency form, like Eq. (1) or Eq. (2), has an equivalent size-frequency one [38]. One should become curious about whether similar equalities hold for the size-frequency cases; (ii) the above considerations suggest to investigate if similar relationships exist for the h-index, distinguishing between h^j, h^p, and h^{jp}, and to draw ıad hoc conclusions; (iii) complex systems do not necessarily lead to find nonlinear laws.

Acknowledgments

This paper is a part of scientific activities in COST Action TD1306 "New Frontiers of Peer Review (PEERE)".

Thanks to H. Bougrine and J. Miskiewicz for discussions on their work [28,29] respectively. Special thanks to J. Miskiewicz for providing unpublished data on polish scientists, with career comments.

Appendix. Other empirical laws: m_a vs. r_M and r_M vs. m_a, with or without outliers

For the theoretical estimation of the numerical values in Eqs. (9)–(13), in particular in order to proceed beyond Eq. (21), it appears that it is useful to find whether a relationship exists between r_M and m_a on average.

The most simple power law fits for m_a vs. r_M, i.e. $m_a \simeq r_M^{\mu}$, give, taking into account all data points,

$$m_a^{(jp)} = 0.392 \quad [r_M^{(jp)}]^{(0.645)}, \quad \text{with } R^2 = 0.893$$
$$m_a^{(j)} = 0.405 \quad [r_M^{(j)}]^{(0.629)}, \quad \text{with } R^2 = 0.978$$
$$m_a^{(p)} = 0.904 \quad [r_M^{(p)}]^{(0.415)}, \quad \text{with } R^2 = 0.760$$

but lead to

$$m_a^{(jp)} = 0.393 \quad [r_M^{(jp)}]^{(0.667)}, \quad \text{with } R^2 = 0.946$$
$$m_a^{(j)} = 0.405 \quad [r_M^{(j)}]^{(0.629)}, \quad \text{with } R^2 = 0.978$$
$$m_a^{(p)} = 0.842 \quad [r_M^{(p)}]^{(0.460)}, \quad \text{with } R^2 = 0.917$$

when the outlier (DG) data point is not taken into account, i.e. roughly speaking $m_a \simeq 0.5\ m_a^{\gamma}$, with $\gamma \simeq 1/2$ or $2/3$. Similarly, the best power law fits for r_M vs. m_a give, when taking into account all data points,

$$r_M^{(jp)} = 8.691 \quad [m_a^{(jp)}]^{(1.175)}, \quad \text{with } R^2 = 0.893$$
$$r_M^{(j)} = 5.013 \quad [m_a^{(j)}]^{(1.483)}, \quad \text{with } R^2 = 0.969$$
$$r_M^{(p)} = 4.261 \quad [m_a^{(p)}]^{(1.507)}, \quad \text{with } R^2 = 0.867$$

but become

$$r_M^{(jp)} = 4.589 \quad [m_a^{(jp)}]^{(1.438)}, \quad \text{with } R^2 = 0.936$$
$$r_M^{(j)} = 5.013 \quad [m_a^{(j)}]^{(1.483)}, \quad \text{with } R^2 = 0.969$$
$$r_M^{(p)} = 2.871 \quad [m_a^{(p)}]^{(1.683)}, \quad \text{with } R^2 = 0.960$$

when the outlier (DG) is not taken into account, i.e. roughly speaking $r_M \simeq 4.5\ m_a^{\beta}$, with $\beta \simeq 3/2$ or $5/3$. The "no (DG)" cases are shown in Figs. 5–6 for illustration of the findings. Note the large R^2 values.

References

[1] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Modern Phys. 81 (2009) 591.
[2] R.S. Mendes, H.V. Ribeiro, F.C.M. Freire, A.A. Tateishi, E.K. Lenzi, Universal patterns in sound amplitudes of songs and music genre, Phys. Rev. E 83 (2011) 017101.
[3] R. da Silva, M.H. Vainstein, S. Gonçalves, F.S.F. Paula, Anomalous diffusion in the evolution of soccer championship scores: Real data, mean-field analysis, and an agent-based model, Phys. Rev. E 88 (2013) 022136.
[4] S.Y. Auyang, Foundations of Complex-Systems, Cambridge University Press, Cambridge, 1998.
[5] R. Albert, A.-L. Barabasi, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47.
[6] N. Boccara, Modeling Complex Systems, Springer-Verlag, New York, 2004.
[7] D. Sornette, Critical Phenomena in Natural Sciences, Springer-Verlag, Berlin, 2006.
[8] A.M. Petersen, H.E. Stanley, S. Succi, Statistical regularities in the rank-citation profile of scientists, Sci. Rep. 1 (2011).
[9] N.K. Ivanov, Z.I. Dimitrova, K.N. Ivanov, Science Dynamics and Scientific Productivity, Vanio Nedkov, Sofia, 2014.
[10] M. Ausloos, A scientometrics law about co-authors and their ranking. The co-author core, Scientometrics 95 (2013) 895–909.
[11] J.E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA 102 (2005) 16569–16572.
[12] J.E. Hirsch, An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship, Scientometrics 85 (2010) 741–754.
[13] R. Rousseau, New developments related to the Hirsch index, Sci. Focus 1 (2006) 23–25.
[14] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, *h*-Index: A review focused in its variants, computation and standardization for different scientific fields, J. Informetr. 3 (2009) 273–289.
[15] G. Laudel, What do we measure by co-authorships? in: M. Davis, C.S. Wilson (Eds.), Proceedings of the 8th International Conference on Scientometrics and Informetrics, Bibliometrics & Informetrics Research Group, Sydney, Australia, 2001, pp. 369–384.
[16] A. Abbasi, J. Altmann, L. Hossain, Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures, J. Informetr. 5 (2011) 594–607.
[17] Chien Hsiang Liao, Hsiuju Rebecca Yen, Quantifying the degree of research collaboration: A comparative study of collaborative measures, J. Informetr. 6 (2012) 27–33.
[18] H. Kretschmer, Cooperation structure, group size and productivity in research groups, Scientometrics 7 (1985) 39–53.
[19] G. Melin, O. Persson, Studying research collaboration using co-authorships, Scientometrics 36 (1996) 363–377.
[20] L.S. Kwok, The White Bull effect: abusive coauthorship and publication parasitism, J. Med. Ethics 31 (2005) 554–556.
[21] Kim A. McDonald, Too many co-authors?, Chron. High. Educ. 41 (33) 35–36.
[22] S. Lee, B. Bozeman, The impact of research collaboration on scientific productivity, Soc. Stud. Sci. 35 (1995) 673–702.
[23] H. Kretschmer, Coauthorship networks of invisible colleges and institutional communities, Scientometrics 30 (1994) 363–369.
[24] A. Zuccala, Modeling the invisible college, J. Am. Soc. Inf. Sci. Technol. 57 (2) (2006) 152–168.
[25] B. Jin, h-Index: An evaluation indicator proposed by scientist, Sci. Focus 1 (1) (2006) 8–9.
[26] B. Jin, The AR-index: complementing the *h*-index, ISSI Newslett. 3 (1) (2007) 6.
[27] B. Jin, L. Liang, R. Rousseau, L. Egghe, The R- and AR-indices: Complementing the *h*-index, Chinese Sci. Bull. 52 (6) (2007) 855–863.

[28] H. Bougrine, Subfield effects on the core of coauthors, Scientometrics 98 (2014) 1047–1064.
[29] J. Miŝkiewicz, Effects of publications in proceedings on the measure of the core size of coauthors, Physica A 392 (2014) 5119–5131.
[30] M. Ausloos, Binary scientific star core size of coauthors, Scientometrics (2014) 1–21.
[31] I. Hellsten, R. Lambiotte, A. Scharnhorst, M. Ausloos, Self-citations networks as traces of scientific careers. In: Ed. by D. Torres-Salinas; H. Moed, Proceedings of the ISSI 2007, 11th International Conf. of the Intern. Society for Scientometrics and Informetrics, CSIC, Madrid, Spain, June 25–27, 2007. Vol. 1, 2007, pp. 361–367.
[32] L. Egghe, R. Rousseau, Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam, 1990.
[33] A.S. Hadi, A.H.M. Rahmatullah Imon, M. Werner, Detection of outliers, WIREs Comput. Stat. 1 (2009) 57–70.
[34] J. Rojo, Heavy-tailed densities, WIREs Comput. Stat. 5 (2013) 30–40.
[35] R.M. Slone, Coauthors contributions to major papers published in the AJR: Frequency of undeserved coauthorship, Am. J. Radiol. 167 (1996) 571–579.
[36] O. Persson, W. Glänzel, R. Danell, Inflationary bibliometric values: The Role of. Scientific collaboration and the need for relative indicators in evaluative studies, Scientometrics 60 (2004) 421–432.
[37] S. Galam, Tailor based allocations for multiple authorship: a fractional *gh*-index, Scientometrics 89 (2011) 365–379.
[38] L. Egghe, Power Laws in the Information Production Process: Lotkaian Informetrics, Elsevier Academic Press, 2005.