Bibliometrics for collaboration works

Paolo Rossi[a], **Alessandro Strumia**[a], **and Riccardo Torre**[b,c]

[a] *Dipartimento di Fisica dell'Università di Pisa, Italy*
[b] *CERN, Theory Division, Geneva, Switzerland*
[c] *INFN, sezione di Genova, Italy*

Abstract

An important issue in bibliometrics is the weighing of co-authorship in the production of scientific collaborations, which are becoming the standard modality of research activity in many disciplines. The problem is especially relevant in the field of high-energy physics, where collaborations reach 3000 authors, but it can no longer be ignored also in other domains, like medicine or biology. We present theoretical and numerical arguments in favour of weighing the individual contributions as $1/N_{\rm aut}^{\alpha}$ where $N_{\rm aut}$ is the number of co-authors. When counting citations we suggest the exponent $\alpha \approx 1$, that corresponds to fractional counting. When counting the number of papers we suggest $\alpha \approx 1/3 - 1/2$, with the former (latter) value more appropriate for larger (smaller) collaborations. We expect and verify that the h index scales as the square root of the average number of co-authors, and define a fractionalized h index that does not scale with collaboration size.

1 Introduction

In many research fields, scientific collaboration has become the standard way of operating, and moreover, due to the increasing complexity of the problems to be faced, the number of scientists with different competences involved in each single collaboration is increasing. In the extreme case of high energy physics numbers have already reached the four-digit level, but in many other domains, like medicine or biology, it is not unusual to find two-digit collaborations.

In the context of bibliometrics this hyper-authorship phenomenon poses a very important question, concerning the individual degree of property that must be assigned to the authors of a common scientific article, both concerning the paper itself and the citations it receives. It is rather clear that attributing the full credit of a paper to each of the authors is mystifying and (if adopted by policymakers) tends to encourage fictitious collaborations, because of the obvious competitive advantage resulting from the much larger number of articles that a collaboration may produce in the same amount of time in comparison with an isolated author. Moreover, also the number of citations received is strongly correlated with the typical dimensions of the collaborations operating in a given field of research.

Fractional counting of papers and citations could be a solution to this issue. Fractional counting has been extensively discussed in the literature. For instance, it has been considered in the context of metrics and rankings by Aksnes, Schneider, and Gunnarsson (2012), Bouyssou and Marchant (2016), Carbone (2011), Egghe (2008), Hooydonk (1997), Leydesdorff and Bornmann (2010, 2011), Leydesdorff and Opthof (2010), Leydesdorff and Shin (2011), Rousseau (2014), Strumia and Torre (2019), and in the context of constructing research networks by Leydesdorff and Park (2017), Perianes-Rodríguez, Waltman, and van Eck (2016). Fractional counting gives an intensive quantity: this means, for example, that the total index of the European Union is the sum of its members, unlike what happens if full counting is adopted (see e.g. the discussions by Gauffriau (2017), Gauffriau, Larsen, Maye, Roulin-Perriard, and von Ins (2008), Strumia and Torre (2019), and Waltman and van Eck (2015)).

However, the choice of fractionally counting papers by attributing a $1/N_{\text{aut}}$ weight to each of the N_{aut} co-authors of a paper would imply a strong penalty for authors beloinging to large collaborations. As we will show in the following, see for instance fig. 3, this can be seen from the strong dependence of the fractionally counted number of papers on the collaboration size, which implies a strong reduction of the fractionally counted number of papers for collaborations with many co-authors. Indeed typically N_{aut} tend to produce less than $N_{\text{pap}} = N_{\text{aut}}$ papers in the same time in which a single author produces a single paper. While it is clear that from a pure research perspective what matters is the scientific impact, that cannot be quantified trough the publication frequency, there are still examples of policymakers that consider the number of papers as a simple relevant indicator and use it for the evaluation of scientists.[1] While full counting of papers penalises authors working in small collaborations (for instance small experiments in fundamental physics), full fractional counting of papers, if adopted by policymakers, would on the contrary discourage large collaborations, which are a necessary endeavour in modern research. This problem needs therefore a non-subjective solution, namely

[1] As an example, the Italian Ministry for Research poses lower thresholds in the publication frequency to access to professorship positions.

finding which compromise between these two extreme choices gives bibliometric indicators that are as independent as possible from the size of collaborations and authors groups.

This general principle will give different answers when applied to different indicators: counting of papers, of citations, h-index, etc.

The aim of the present paper is to find fractional counting algorithms that do not scale with collaboration size. This has a two-fold advantage: on the one hand it ensures that neither authors belonging to large collaborations, nor single authors nor authors working in small groups are favoured or disfavoured, when bibliometric indicators are is used in their evaluation, for their choice of carrying out their research in large versus small groups. On the other hand this allows to better quantify and qualify the bibliometric output of large collaborations in comparison with small groups of authors (or even single authors).

The bibliometric literature documents that collaboration papers tend to have higher impact than single-authored papers. Beaver (1986) studied the field of physics, finding that co-authored research tends to be of higher quality than solo research. Bordons, Garcia-Jover, and Barrigon (1993) studied Spanish publications in pharmacology and pharmacy finding that internationally co-authored documents have higher impact than the remaining collaborative documents or non-collaborative ones. Avkiran (2013) found that collaboration leads to articles of higher impact in finance, up to 4 collaborators. Gazni and Didegah (2011) found a significant positive correlation between the number of authors and the number of citations in Harvard publications. Hsu and Huang (2011) considered 90k articles in natural sciences, finding that the average number of citations scales as $N_{\rm aut}^{1/3}$ (data extend up to about 10 co-authors), up to wide fluctuations. Lee and Bozeman (2005) found that the number of peer-reviewed journal papers is strongly and significantly associated with the number of collaborators, unlike the number of fractionally-counted papers. Katz and Hicks (1997) studied how the average number of citations per paper varies with different types of collaborations. See also the works of van Raan (1997), Sooryamoorthy (2009) and Birnholtz (2006) for additional studies.

General theoretical arguments concerning scale-free systems suggest that the scientific productivity of collaborations and the corresponding frequency distribution of citations should show some, at least approximate, power law dependence on $N_{\rm aut}$. Empirical evidence appears to support these arguments. Finding the most appropriate exponents for these scaling laws would offer the possibility of weighing the production of collaborations in the bibliometric estimate of the (quantitative) value of their results in such a way as to discourage adaptive and opportunistic behaviours while encouraging more appropriate practices in the indication of co-authorship.

In section 2 we develop and present some theoretical arguments in favour of weighing the individual contributions to a single paper as $1/N_{\rm aut}^{\alpha}$, where $\alpha \approx 1/3 - 1/2$, with the former (latter) value more appropriate for larger (smaller) collaborations. When counting overall citations we suggest the exponent $\alpha = 1$, corresponding to fractional counting. By combining the two above arguments, in section 2.4 we define an h index that does not scale with collaboration size.

In section 3 we analyze empirical data concerning a very large number of collaborations active in fundamental physics, where the range of available values of $N_{\rm aut}$ allows for sufficiently convincing estimates of the exponents describing the dependence on $N_{\rm aut}$ of the total number of papers and of the mean and total number of citations.

Finally, we summarize and draw our conclusions in section 4.

2 A theoretical approach

2.1 Scaling

The behaviour of collaborations with $N_{\rm aut}$ authors can be viewed as a scale-free phenomenon for a wide range of values of $N_{\rm aut}$. Any upper limit on $N_{\rm aut}$ would be sufficiently large to exclude any sensible effect on the equilibrium distributions in the range of values we are interested to explore (3 to 4 orders of magnitude).

We therefore expect that the various indices N_I that characterise bibliometric outputs of collaborations are distributed at equilibrium following a power-law behaviour, which we parametrise as follows

$$\langle N_I \rangle = C_I N_{\rm aut}^{p_I} \tag{1}$$

where C_I and the powers p_I are constants. For example this applies to the number of papers produced (in a definite amount of time) by a scientific collaboration $N_{\rm pap} = C_{\rm pap} N_{\rm aut}^{p_{\rm pap}}$ and to the average number of citations per paper $N_{\rm cit} = C_{\rm cit} N_{\rm aut}^{p_{\rm cit}}$. The total number of citations $N_{\rm totcit}$ received by the papers of a collaboration then scales as

$$N_{\rm totcit} = N_{\rm cit} N_{\rm pap} = C_{\rm cit} C_{\rm pap} N_{\rm aut}^{p_{\rm totcit}}, \qquad p_{\rm totcit} = p_{\rm cit} + p_{\rm pap}. \tag{2}$$

2.2 Scaling of the total number of citations

Assuming a collective rational behaviour, and that on average the number of citations received by scientific papers may be considered as a reasonable proxy for their quality, we might expect that individual and collective choices would lead at equilibrium to

$$p_{\rm totcit} \approx 1, \tag{3}$$

namely that the total number of citations received by collaborations scales, on average, with the number of members. This expresses the fact that the (average) value of work made by $N_{\rm aut}$ scientists should approximately be equal to $N_{\rm aut}$ times the work made by a single scientist. A lower power $p_{\rm totcit}$ would arise in the presence of gift authorships, namely of authors who sign papers without substantially contributing.

This means that the total number of citations is not a fair indicator for authors, as it grows with the number of co-authors. Similarly, the h-index introduced by Hirsch (2005), being on average proportional to square root of the number of citations, grows on average as the square root of the number of co-authors.

A bibliometric index which does not overestimate nor underestimate individual contributions to a collaboration is then the number of fractionally-counted citations $N_{\rm fcit}$ received by each author. This means that a fraction f_A of each paper is attributed to each co-author A such that the fractions f_A sum up to unity.[2] On average $N_{\rm fcit}$ scales with power index $p_{\rm fcit} = p_{\rm totcit} - 1 \approx 0$, showing that $N_{\rm fcit}$ is a scale-invariant quantity that (unlike the number of citations) cannot be arbitrarily inflated grouping authors.

2.3 Scaling of total number of papers

In order to implement these concepts into actual bibliometric indices for the total number of papers or for the average number of citations, we must offer arguments in favour of explicit values for the exponents $p_{\rm pap}$ and $p_{\rm cit}$.[3]

We present here a simple "theoretical" argument. As the goal of collaborations is achieving more than what single authors can achieve, we expect $p_{\rm cit} > 0$. Assuming rational behaviour in the formation of collaborations one may expect that (at least for not-too-big groups), individual competence of partners be as far as possible complementary, and therefore "orthogonal" in some abstract N-dimensional "space of competencies". We may therefore regard the qualitative output of a collaboration as the vector sum of N orthogonal vectors.

The limit where all authors have "orthogonal" competencies corresponds to $N = N_{\rm aut}$: then the length of such vector scales as $N_{\rm aut}^{1/2}$.[4] This corresponds to

$$p_{\rm pap} = 1/2, \qquad p_{\rm cit} = 1/2. \tag{4}$$

Sometimes, more than one collaborator is needed to fulfil a needed competency: realistic collaborations organise in $N_{\rm sub} \le N_{\rm aut}$ sub-collaborations that work on "orthogonal" topics. We assume the number of sub-collaborations satisfies the scaling law

$$N_{\rm sub} = N_{\rm aut}^s \qquad \text{with exponent } s \le 1. \tag{5}$$

[2] We do not address the relative assignment of credit. In some fields the contribution of different authors is reflected by their order, with special recognition given to first and last authors. Various proposals have been put forward to encode the relative credit in the fractions f_A (see, e.g. Kosmulski (2012) and Waltman (2015)). In some other fields authors are sorted alphabetically, giving no information about who contributed more. This happens in most papers in our data-base, so that we will assume a common f_A equal to the inverse of the number of authors.

[3] Some authors think that counting publications has no bibliometric interest, with citations being the only relevant quantity to be measured. From such a perspective, it remains nevertheless interesting to know how collaborations tend to split their bibliometric output within their publications. For example one might observe a gap in citation output between (groups of) authors and want to understand if it mostly arises from publication intensity. As another example, one might have a partial data-base (e.g. limited to some area) that only allows to reliably compute publication intensity. Furthermore, the publication indicator is available immediately, while the citation indicator becomes more significant after some years as citations accumulate. As a matter of fact, while publication intensity is a less significant indicator, it remains used or at least mentioned because of its simplicity. At the opposite extremum other authors think that citations can be significantly distorted by social biases, and view publications numbers as a more objective bibliometric indicator. In order to obtain a result that does not scale with collaboration size, one should then fractionalize papers according to the appropriate power for papers.

[4] Assuming that authors have different skills, the average length squared of such vector scales as $N_{\rm aut}$.

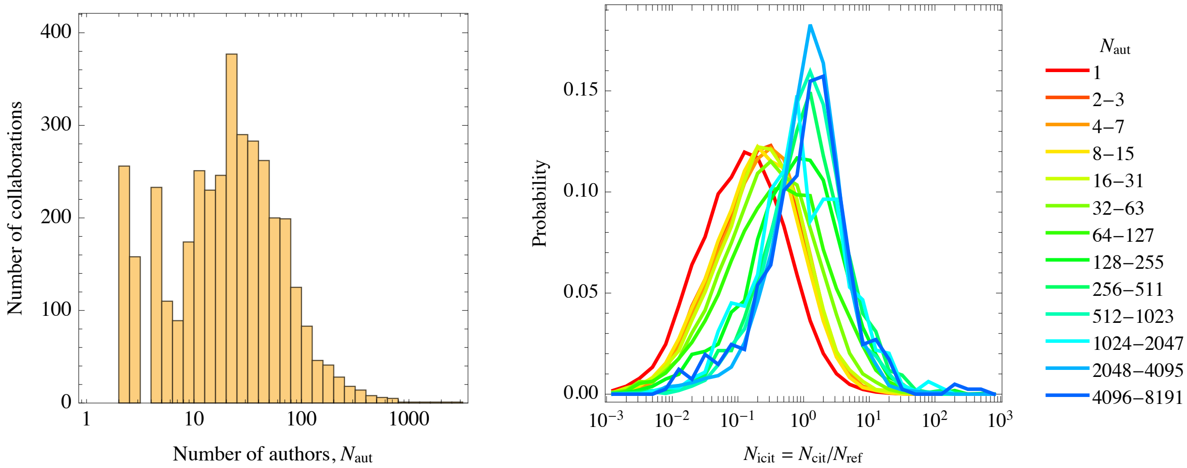

Figure 1: **Left**: Total number of collaborations listed in the InSpire database with the number of authors shown on the horizontal axis. **Right**: distribution of the number of individual citations (citations divided by the number of references of the citing papers) received by papers with the indicated number of authors.

Then, the average number of citations of each paper scales as the square root of the number of "orthogonal" competencies $N = N_{\rm sub}$:

$$N_{\rm cit} \propto \sqrt{N} \propto N_{\rm aut}^{s/2}. \qquad (6)$$

Assuming again an optimal distribution of resources, $N_{\rm totcit}$ is expected to scale as $N_{\rm aut}$, and thereby the number of papers is expected to scale as

$$N_{\rm pap} \propto N_{\rm aut}^{1-s/2}. \qquad (7)$$

It is reasonable to assume that the number of papers scales as N, the number of topics about which the collaboration has competencies. This leads to $s = 1 - s/2$, solved by $s = 2/3$, and thereby to

$$p_{\rm pap} = 2/3, \qquad p_{\rm cit} = 1/3. \qquad (8)$$

A weaker growth of the number of papers with N leads to smaller $p_{\rm pap}$.

2.4 Scaling of the h index

The h index (defined by Hirsch (2005) as the number of papers that received more than h citations), provides extra information on the distribution of the number of citations, favouring authors that produced many highly-cited papers with respect to authors that produced many poorly cited papers plus a small number of top-cited papers.

Theoretical arguments (see e.g. Yong (2014)) and evidence from data analysis (see e.g. Mannella and Rossi (2013) and Strumia and Torre (2019)) indicate that the h index is strongly correlated to the square root of the total number of citations received by an author:

$$h \approx \alpha\, N_{\rm totcit}^{0.5}. \tag{9}$$

where the theoretical prediction from Yong (2014) is $\alpha \approx 0.54$ and the phenomenological result obtained from the data of about 1400 Italian physicists is $\alpha \approx 0.53$ (Mannella and Rossi, 2013).

Like the number of citations, the h index is affected by the collaboration size, being higher for authors with more collaborators. Assuming for the h index the above mentioned scaling as a function of $N_{\rm totcit}$, and recalling our prediction $p_{\rm totcit} \approx 1$, we thereby expect that the h-index should scale approximately as the square root of the (average) number of authors:

$$h \propto N_{\rm aut}^{0.5} \tag{10}$$

independently of the specific values taken by $p_{\rm pap}$ and $p_{\rm cit}$, as long as they satisfy $p_{\rm cit} + p_{\rm pap} \approx 1$.

3 Data about collaborations in fundamental physics

In this section we present bibliometric data in fundamental physics, that offer support for

$$p_{\rm pap} \approx 0.5 - 0.6, \qquad p_{\rm cit} \approx 0.4 - 0.5, \qquad p_{\rm totcit} \approx 1, \qquad p_{\rm fcit} \approx 0. \tag{11}$$

We use the InSpire[5] database that gives a picture of fundamental physics world-wide from ~ 1970 to end 2019: 1.34 millions of scientific papers, 32 millions of references, 75 thousands of identified authors. Fundamental physics contains large collaborations, up to 3000 authors that produced 6000 publications. Adopting full counting these are counted as 3000×6000 publications, dominating the whole database and producing bibliometric indices uncorrelated to human evaluations of scientific merit. Fundamental physics thereby is a good sample to study how the bibliometric outputs of collaborations scale with the number of collaborators. We will show data for two different kinds of collaborations:

1. **Official collaborations**. We consider the 5965 (mostly experimental) collaborations listed in the InSpire database. Each collaboration produced a certain number $N_{\rm pap}$ of papers, roughly written with the same group of $N_{\rm aut}$ authors. The left panel of Figure 1 gives some demographic information, that is the distribution of the number of authors in collaborations and the distribution of fractionally counted citations for different collaboration sizes. In the following, when showing results for official collaborations, we indicate collaborations as dots in a scatter plot, with the main collaborations indicated by their names. Furthermore, we show the mean (median) as a red (magenta) curve and a blue dotted line highlighting the scaling with the number of authors.[6]

[5] High-Energy Physics Literature InSpire Database (https://inspirehep.net).

[6] A few collaborations varied significantly their number of authors. We define the number of authors of a collaboration by averaging the number of authors of all its papers, with weights proportional to their number of citations. This procedure assigns minor weight to proceedings written by one or few authors and to papers written by earlier incomplete phases of the collaboration.

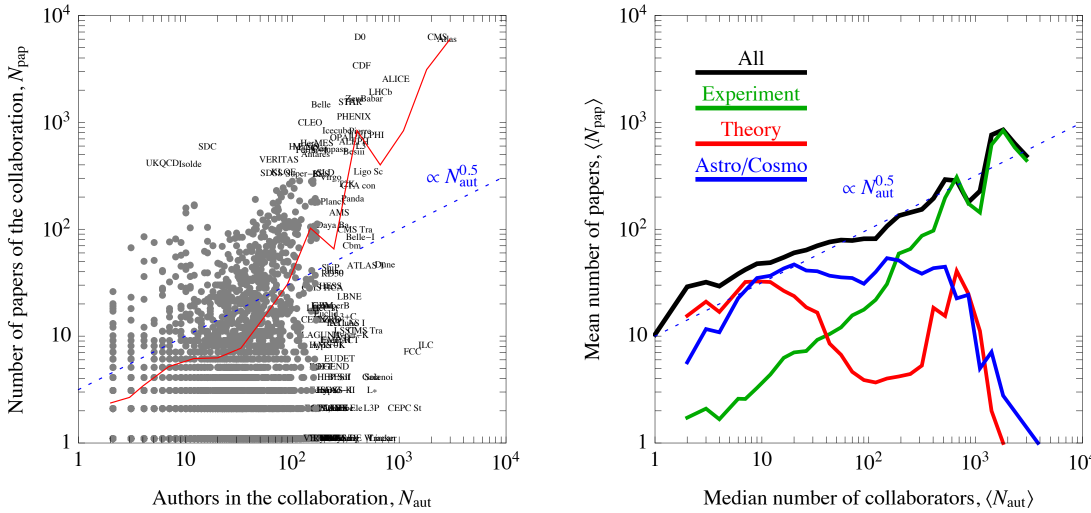

Figure 2: Number of papers versus number of collaborators.

2. **Occasional collaborations**. Many more multi-authored papers have been written by collaborations that form for one or few papers. To study them we proceed as follows. For each author in the InSpire database we compute the average number of authors of his/her papers, $\langle N_{aut} \rangle \geq 1$, as well as his/her bibliometric indices (number of papers, of citations, etc). In view of the large number of authors, in the following, when presenting results on occasional collaborations, we avoid showing scatter plots and only show averages. Moreover, results are shown separately within the main topics of fundamental physics: experiment, theory, astro/cosmo. The first category includes all papers in the hep-ex (high-energy experiments) and nucl-ex (nuclear experiments) category of arXiv. The latter category includes papers in astro-ph, which contains astrophysics and cosmology. Theoretical papers are those appeared in hep-ph (high-energy phenomenology), hep-th (high-energy theory), hep-lat (lattice), nucl-th (nuclear theory), gr-qc (general relativity and quantum cosmology).

3.1 Scaling of the number of papers

Figure 2 shows that the number of papers produced by official (left panel) or occasional (right panel) collaborations scales with the number of authors as

$$N_{pap} \propto N_{aut}^{0.5-0.6}. \tag{12}$$

In the right panel, theoretical papers with many authors fall below the scaling. These are rare outliers: almost all papers in theoretical categories have few authors. Theoretical papers with many authors mostly are collections of separate contributions grouped together, rather than big collaborations.

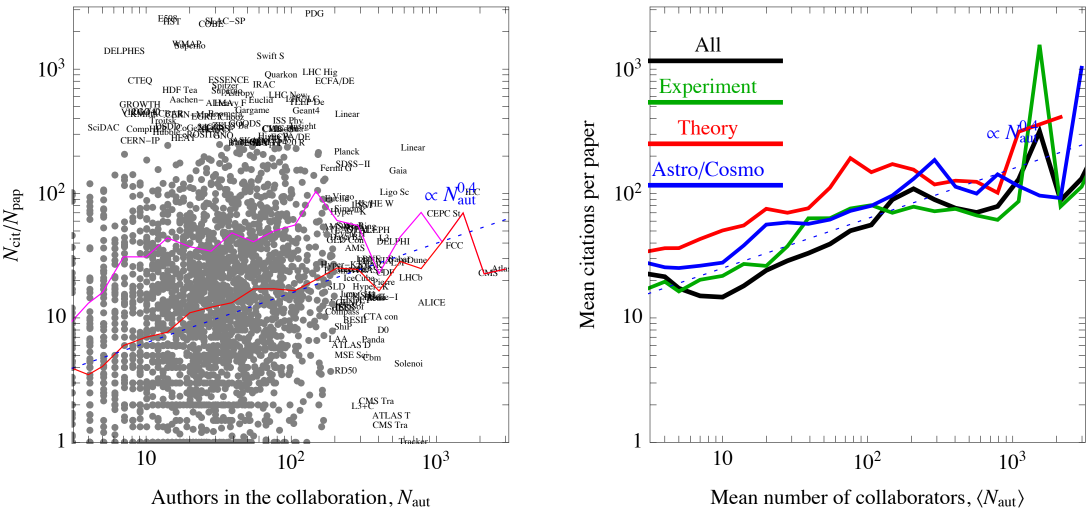

Figure 3: Mean number of citations per paper versus number of collaborators.

3.2 Scaling of the mean number of citations

Figure 3 shows that the mean number of citations received by papers written by an official collaboration (left panel) or by an author (right panel) roughly scales with the average number of co-authors as

$$N_{cit} = \frac{N_{totcit}}{N_{pap}} \propto N_{aut}^{0.4-0.5}. \tag{13}$$

This result is in reasonable agreement with Hsu and Huang (2011), who found, in a much smaller sample, a power $\approx 1/3$ up to $N_{aut} \sim 10$.

3.3 Scaling of the total number of citations

Figure 4 shows that the total number of citations received by an official collaboration (left panel) or author (right panel) grows roughly linearly with the average number of co-authors:

$$N_{totcit} \propto N_{aut}^{1}. \tag{14}$$

This is expected by combining the two previous scalings: the total number of citations of a collaboration can be decomposed as the product of the number of papers written by the collaboration, times the average number of citations received per paper: these factors roughly scale as $N_{aut}^{0.5-0.6} \times N_{aut}^{0.4-0.5}$.

Figure 5 shows that the total number of fractionally-counted citations $N_{fcit} = \sum_p N_{pcit}/N_{paut}$ received by papers p written by an official collaboration (left panel) or author (middle panel) is roughy independent of the average number of co-authors. A similar result holds for a related quantity, "individual citations", defined as fractionally counted citations divided by the number of references of the citing papers:

$$N_{fcit}, N_{icit} \propto N_{aut}^{0}. \tag{15}$$

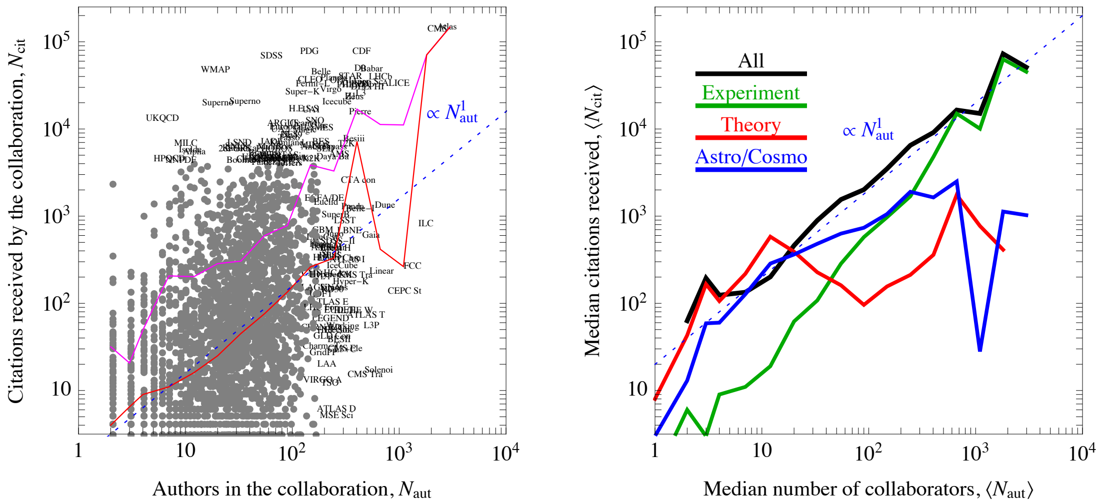

Figure 4: Number of citations versus number of collaborators.

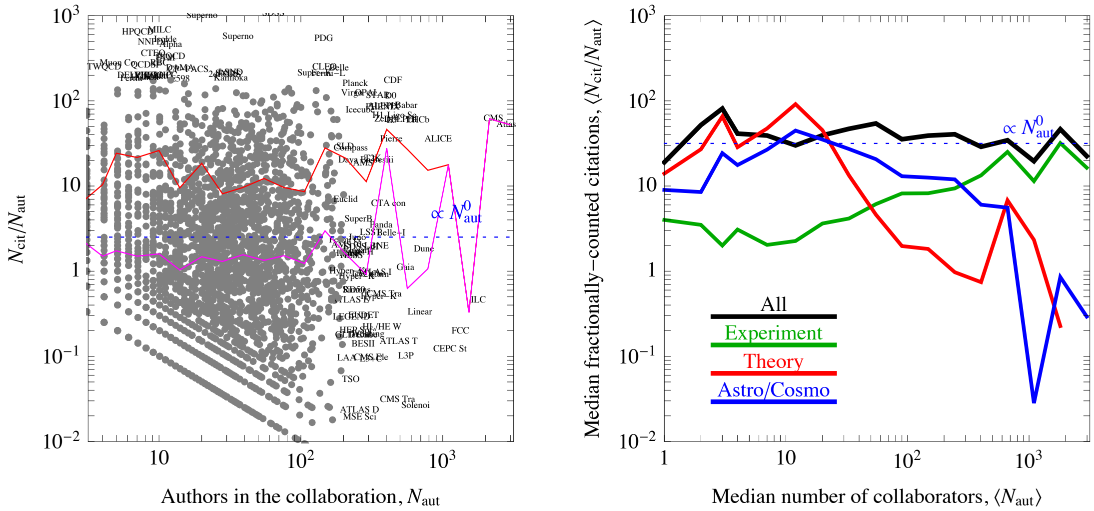

Figure 5: Total number of fractionally-counted citations versus number of collaborators.

This means that fractionally-counted citations or individual citations neither reward nor penalise working in big collaborations, while citations reward authors who prefer working in big collaborations.

3.4 Scaling of the total number of field-normalized citations

When analysing different fields it is usually meaningful to correct for their different publication intensities. As different fields also show different collaboration patterns, this opens the issue about how to properly account for the two aspects in a combined way. A simple general answer is obtained by counting citations divided by the number of references of the citing

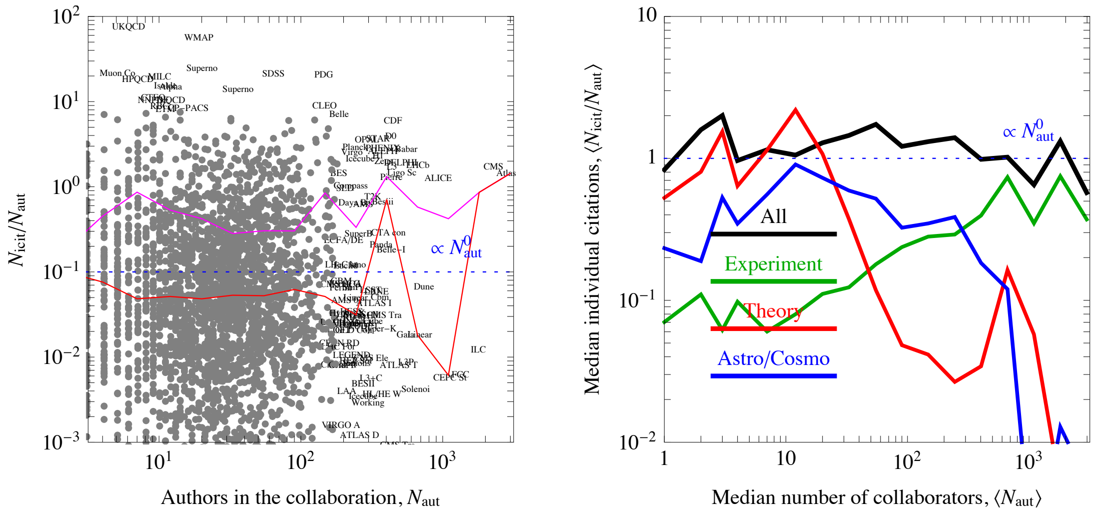

Figure 6: Total number of individual citations (fractionally-counted citations divided by references of citing papers) versus number of collaborators.

paper, namely $N_{\rm icit} \equiv N_{\rm cit}/N_{\rm aut}^{p_{\rm totcit}} N_{\rm ref}$ (for a precise definition see Strumia and Torre (2019) where this indicator is dubbed "individual citations"). This quantity is similar to citations (so that $p_{\rm totcit} \approx 1$) and it automatically provides a field-independent indicator, without having to identify fields. Indeed, within any hypothetical closed field (with no citations to or from other fields) this indicator satisfies the sum rule $\sum N_{\rm icit} = N_{\rm pap}$, up to recent papers that will receive citations in the future. In words, papers in sectors with higher publication intensity tend to receive more citations and thereby also tend to have more references. One can thereby use references to factor out publication intensity (see e.g. Zitt and Small (2008) and Waltman (2015) for a review).

Figure 6 shows this field-normalised indicator applied to fundamental physics, showing that it exhibits the desired negligible scaling with collaboration size.

3.5 Scaling of the (fractionalized) Hirsch h index

Looking at data, the top-left panel of fig. 7 shows that the h index scales with the number of authors as

$$h \propto N_{\rm aut}^{0.5-0.6} \tag{16}$$

consistently with our expectation. In order to avoid this scaling, various authors defined modified h indices with fractionalized counting. As the h index combines information on the number of citations and of papers, one can fractionalize with respect to citations and/or with respect to papers. The first option was considered by Egghe (2008), who defined an index (here called $h_{\rm cit}$) equal to the number of papers that received more than $h_{\rm cit}$ fractionalized citations. The second option was considered by Schreiber (2008), who defined an index (here called $h_{\rm pap}$) equal to the number of fractionally-counted papers that received more than $h_{\rm pap}$ citations. Looking at data, the lower row of fig. 7 shows that the these two possibilities are only partially successful,

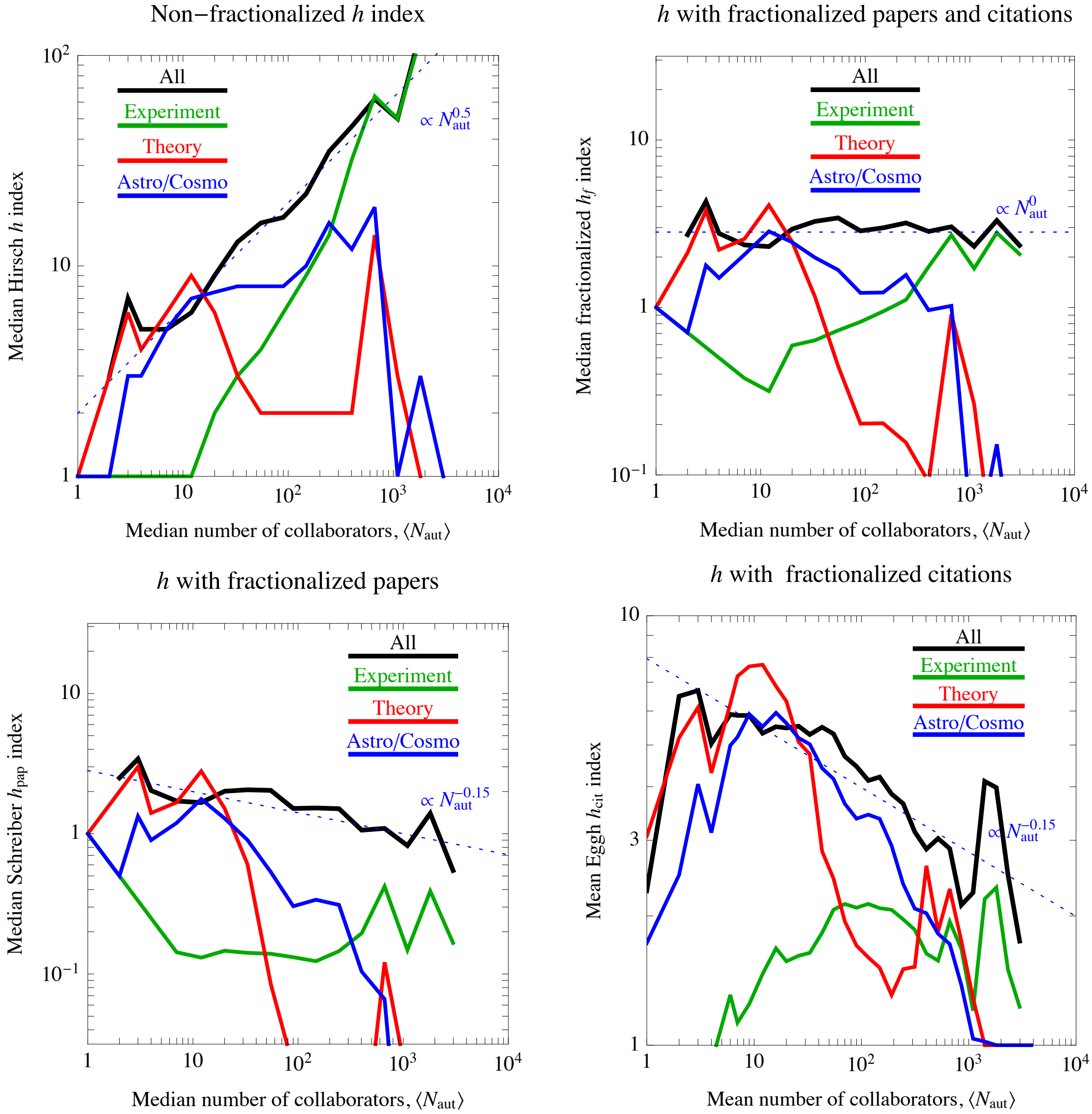

Figure 7: **Top-left:** the h-index scales as the square root of the average number of collaborators. **Top-right**: the h-index fractionalized as suggested in eq. (18) does not scale with the average number of collaborators. **Bottom**: the h-index fractionalized with respect to papers (left) or to citations (right) shows a reduced residual scaling with the average number of collaborators.

as they both still exhibit a milder scaling with the average number of co-authors

$$h_{\rm cit}, h_{\rm pap} \propto N_{\rm aut}^{-0.15}. \tag{17}$$

In order to understand the reason of this residual scaling with the average number of co-authors, we recall that larger collaborations tend to produce more papers as well as papers that get more cited. The h index fractionalized with respect to papers would exhibit the desired scale-invariant behaviour if the scalings were $N_{\rm pap} \propto N_{\rm aut}^{p_{\rm pap}}$ and $N_{\rm cit} \propto N_{\rm aut}^{p_{\rm cit}}$ with $p_{\rm pap} = 1$ and

$p_{\rm cit} = 0$. Similarly, the h index fractionalized with respect to citations would exhibit the desired scale-invariant behaviour if $p_{\rm pap} = 0$ and $p_{\rm cit} = 1$.

Instead, data show $p_{\rm pap} \approx p_{\rm cit} \approx 1/2$ (see eq. (12) and eq. (13)), in agreement with our model values of eq. (4). Thereby a fractionalized h-index h_f that exhibits the desired scale-invariant behaviour is obtained by partially fractionalizing with respect to both papers and citations. To compute our h_f one needs to sort papers according to citations partially fractionalized as $c = N_{\rm cit}/\sqrt{N_{\rm aut}}$, and summing authorships partially fractionalized as $1/\sqrt{N_{\rm aut}}$ until h_f is larger than the fractionalized c. In formulæ

$$h_f = \sum_{h_f < N_{\rm cit}/\sqrt{N_{\rm aut}}} \frac{1}{\sqrt{N_{\rm aut}}} \tag{18}$$

where the sum runs over all papers of the considered author. The top-right panel of fig. 7 shows that h_f exhibits the expected scale-invariant behaviour, and can thereby be used to quantify the scientific output of authors in a way that does not depend on their average number of collaborators.

Another simpler index that achieves the same independence is the number of fractionally counted citations $N_{\rm fcit}$ plotted in fig. 5. The h_f index, analogously to the h index, provides extra information on the distribution of the number of citations, and has the same intuitive meaning as the Schreiber h_m index, that here we called $h_{\rm pap}$.

4 Summary and conclusions

In this paper we studied the bibliometric output of collaborations. Our study was motivated by the present situation in fundamental physics, where collaborations can be so large that accounting for their size has a huge bibliometric impact. Nevertheless, our results also apply to fields with smaller collaborations, up to the quantitative difference that the factors we considered have less numerical relevance.

Clearly, some partial fractional counting of bibliometric quantities is the solution to the problem. However, unless the correct amount of fractional counting appropriate to any given bibliometric indicator is used, the index will favour or disfavour scientists that choose to carry out their research in small or large groups. This is a particularly relevant matter, given that there are nowadays several policymakers that implement bibliometric evaluation in their funding and hiring criteria.

The field of fundamental research represents the natural playground to study bibliometrics for collaborations, both in terms of the large available dataset provided by InSpire and in terms of the large variability in collaboration size, that is by far the largest in the public research domain, reaching thousands of authors. To our knowledge there are no other studies on this topic carried out with such a dataset, so that our results cannot be confronted with previous studies.

Having to deal with $N_{\rm aut} \gg 1$ authors, we started in section 2 with general theoretical considerations on the behaviour of scientific research that lead us to a series of assumptions and to the subsequent formulation of hypotheses on the scaling of bibliometrics indicators with the number of authors. The crucial hypotheses are that bibliometric quantities for collaboration work exhibit a power-law behaviour with the number of authors in the collaboration, and that, in a situation of "equilibrium" the scaling of the total number of citations received by a collaboration, summing up all its papers, scales linearly with the number of authors. Moreover, in the hypothetical situation in which all authors contribute with fully "orthogonal" competences, the scaling of the number of papers and of the number of citations per paper are both equal to 1/2. We therefore formulated a theoretical model for bibliometrics of collaborations based on the aforementioned considerations and assumptions. As any model, it needs to be confronted with data to extract unknowns, that, in our case are the scaling exponents of the power-law dependence of the number of papers $p_{\rm pap}$, the number of citations per paper $p_{\rm cit}$, the total number of citations $p_{\rm totcit}$, and the number of fractionally counted citations $p_{\rm fcit}$.

In section 3 we computed all the quantities relevant for our model and extracted the unknowns from the data in the InSpire dataset. On the one hand we observed in data the expected approximate power-law scaling. On the other hand, we were able to estimate the desired exponents, observing $p_{\rm pap} \approx 0.5 - 0.6$, $p_{\rm cit} \approx 0.4 - 0.5$, $p_{\rm totcit} \approx 1$, and $p_{\rm fcit} \approx 0$.

Incidentally, given the relatively common use of the h index, despite its well known correlation with the number of citations, we also predicted and evaluated the scaling with the number of authors of the h index, $h \propto N_{\rm aut}^{0.5-0.6}$. Furthermore, we defined a modified h index (see eq. (18)) that roughly does not scale with the number of authors. We confronted our result with different modified h indices proposed in the literature, which did not solve the issue of dependence on the number of authors.

Our results apply to mean (or median) quantities. Before concluding, a comment on the full distributions, or at least on their variances is in order. In the right panel of Figure 1 we show the distributions of individual citations received by all papers in our database, splitting them according to their number of authors. We see that, as already suggested in the literature, papers with more authors are more cited. We also see that distributions have large variabilities: the distributions are approximatively log-normal with log-scale means that scale as $\langle N_{\rm cit}\rangle \propto N_{\rm aut}^{0.5}$ and with log-scale widths that remain approximatively constant. This behaviour is obtained from our initial theoretical considerations adding one extra assumption: that collaborations tend to equalise the total amount of skill within each competence, such that the distribution in $N_{\rm cit}$ of a collaboration is simply obtained rescaling the distribution of single-author papers. The bibliometric output of collaborations formed as random groups of authors would instead show larger variabilities.

References

Aksnes, D. W., Schneider, J. W., and Gunnarsson, M. (2012). Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods. *J. Informetrics*, *6*(1), 36–43. (Semantic Scholar) doi: [link]

Avkiran, N. K. (2013). An empirical investigation of the influence of collaboration in finance on article impact. *Scientometrics*, *95*(3), 911-925. doi: [link]

Beaver, D. B. (1986). Collaboration and teamwork in physics. *Czech. J. Phys. B*, *36*(1), 14-18. (ADS) doi: [link]

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *JASIST*, *57*(13), 1758-1770. (ACM DL) doi: [link]

Bordons, M., Garcia-Jover, F., and Barrigon, S. (1993). Is collaboration improving research visibility? Spanish scientific output in pharmacology and pharmacy. *Research Evaluation*, *3*(1), 19-24. doi: [link]

Bouyssou, D. and Marchant, T. (2016). Ranking authors using fractional counting of citations: An axiomatic approach. *J. Informetrics*, *10*(1), 183–199. (Semantic Scholar) doi: [link]

Carbone, V. (2011). Fractional counting of authorship to quantify scientific research output. *arXiv*, *1106.0114*. (Semantic Scholar)

Egghe, L. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. *JASIST*, *59*(10), 1608–1616. (Semantic Scholar) doi: [link]

Gauffriau, M. (2017). A categorization of arguments for counting methods for publication and citation indicators. *J. Informetrics*, *11*(3), 672–684. (Semantic Scholar) doi: [link]

Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., and von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics*, *77*(1), 147–176. (Semantic Scholar) doi: [link]

Gazni, A. and Didegah, F. (2011). Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications. *Scientometrics*, *87*(2), 251-265. (Semantic Scholar) doi: [link]

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc. Nat. Acad. Sci.*, *46*. (Semantic Scholar) doi: [link]

Hooydonk, G. V. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *JASIS*, *48*(10), 944–945. (Semantic Scholar) doi: [link]

Hsu, J. and Huang, D. (2011). Correlation between impact and collaboration. *Scientometrics*, *86*(2), 317-324. doi: [link]

Katz, J. S. and Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, *40*(3), 541-554. (Semantic Scholar) doi: [link]

Kosmulski, M. (2012). The order in the lists of authors in multi-author papers revisited. *J. Informetrics*, *6*(4), 639–644. (Semantic Scholar) doi: [link]

Lee, S. and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673-702. doi: [link]

Leydesdorff, L. and Bornmann, L. (2010). How fractional counting affects the impact factor: Steps towards field-independent classifications of scholarly journals and literature. *arXiv*, *1007.4749*. (Semantic Scholar)

Leydesdorff, L. and Bornmann, L. (2011). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *JASIST*, *62*(2), 217–229. (Semantic Scholar) doi: [link]

Leydesdorff, L. and Opthof, T. (2010). Normalization at the field level: fractional counting of citations. *J. Informetrics*, *4*, 644-646. (Semantic Scholar)

Leydesdorff, L. and Park, H. W. (2017). Full and fractional counting in bibliometric networks. *J. Informetrics*, *11*(1), 117–120. (Semantic Scholar) doi: [link]

Leydesdorff, L. and Shin, J. C. (2011). How to evaluate universities in terms of their relative citation impacts: Fractional counting of citations and the normalization of differences among disciplines. *JASIST*, *62*(6), 1146–1155. (Semantic Scholar) doi: [link]

Mannella, R. and Rossi, P. (2013). On the time dependence of the *h*-index. *J. Informetrics*, *7*(1), 176–182. (Semantic Scholar) doi: [link]

Perianes-Rodríguez, A., Waltman, L., and van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *J. Informetrics*, *10*(4), 1178–1195. (Semantic Scholar) doi: [link]

Rossi, P., Strumia, A., and Torre, R. (2019). Bibliometrics for collaboration works. In *Proceedings of the 17th International Conference on Scientometrics and Informetrics ISSI 2019, Rome, 2-5 September 2019* (Vol. I, p. 975-983). (Semantic Scholar)

Rousseau, R. (2014). A note on the interpolated or real-valued h-index with a generalization for fractional counting. *Aslib J. Inf. Manag.*, *66*(1), 2–12. (Semantic Scholar) doi: [link]

Schreiber, M. (2008, jul). A modification of the h-index: the hm-index accounts for multi-authored manuscripts. *Journal of Informetrics*, *2*(3), 211–216. (Semantic Scholar) doi: [link]

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, *81*, 177-193. (Semantic Scholar) doi: [link]

Strumia, A. and Torre, R. (2019). Biblioranking fundamental physics. *J. Informetrics*. (Semantic Scholar)

van Raan, A. F. J. (1997). Science as an international enterprise. *Science and Public Policy*, *24*(5), 290-300. doi: [link]

Waltman, L. (2015). A review of the literature on citation impact indicators. *CoRR*, *abs/1507.02099*. Retrieved from `http://arxiv.org/abs/1507.02099` (Semantic Scholar)

Waltman, L. and van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *J. Informetrics*, *9*(4), 872–894. (Semantic Scholar) doi: [link]

Yong, A. (2014). Critique of hirsch's citation index: a combinatorial fermi problem. *Notices of the AMS*, *61*(9), 1040–1050. (Semantic Scholar) doi: [link]

Zitt, M. and Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and technology*, *59*(11), 1856–1860. (Semantic Scholar) doi: [link]