Biblioranking fundamental physics

Alessandro Strumia\textsuperscript{a,b,c}, Riccardo Torre\textsuperscript{a,d}

\textsuperscript{a} CERN, Theory Division, Geneva, Switzerland
\textsuperscript{b} Dipartimento di Fisica dell’Università di Pisa, Italy
\textsuperscript{c} INFN, sezione di Pisa, Italy
\textsuperscript{d} INFN, sezione di Genova, Italy

Abstract

We propose measures of the impact of research that improve on existing ones such as counting of number of papers, citations and $h$-index. Since different papers and different fields have largely different average number of co-authors and of references we replace citations with individual citations, shared among co-authors. Next, we improve on citation counting applying the PageRank algorithm to citations among papers. Being time-ordered, this reduces to a weighted counting of citation descendants that we call PaperRank. Similarly, we compute an AuthorRank applying the PageRank algorithm to citations among authors. These metrics quantify the impact of an author or paper taking into account the impact of those authors that cite it. Finally, we show how self- and circular- citations can be eliminated by defining a closed market of citation-coins. We apply these metrics to the InSpire database that covers fundamental physics, ranking papers, authors, journals, institutes, towns, countries, continents, genders, for all-time and in recent time periods.
1 Introduction

Bibliometrics can be a useful tool for evaluating research: it provides simple, quick, first objective measures of the impact of papers and authors and is increasingly being considered a useful (although incomplete) evaluation criterion in postdoc/faculty recruitments, fundings, and grant awards [1–3]. Goodhart’s law states that: “when a measure becomes a target, it ceases to be a good measure”. This is happening with the most common bibliometric indicators. For instance, in the fundamental physics community that we consider in this paper, all common measures of the impact such as counting of number of papers, citations and Hirsch’s h-index [4], are inflating, making it harder to identify the real impact of research, especially for the most recent literature. The more papers one writes and the more citations these papers get, the biggest bibliometric estimators become: it does not matter if these citations close in
certain loops and/or remain confined in sub-fields, or whether the paper has been written by a single author or in collaboration with thousands of people.

We introduce new metrics and compare them with the existing ones, showing how they address the issues mentioned above. We apply them to the INSPIRE bibliographic database [5–9],¹ that covers fundamental physics literature after ≈ 1970.² The metrics, both the usual ones and the new ones that we introduce, can measure the impact of papers, \( p, p', \ldots \), of authors, \( A, A', \ldots \), and of groups. They are defined as follows:

1. **Number of papers**

   The most naive metric consists in counting the number of papers \( N_{pap}^A = \sum_{p \in A} 1 \) written by a given author \( A \). This metric rewards the most prolific authors.

2. **Number of citations**

   The most used metric consists in counting the number of citations \( N_{cit}^p \) received by a paper \( p \). An author \( A \) is then evaluated summing the number of citations \( N_{cit}^A \) received by its papers. In formulæ:

   \[
   N_{cit}^p = \sum_{p' \to p} 1, \quad N_{cit}^A = \sum_{p \in A} N_{cit}^p, \tag{1}
   \]

   where the first sum runs over all papers \( p' \) that cite \( p \), and the second sum over all papers \( p \) of author \( A \).

3. **h-index**

   The \( h \)-index is defined as the maximum \( h \) such that \( h \) papers have at least \( h \) citations each. In formulæ, assuming that all papers of author \( A \) are sorted in decreasing order of number of citations \( N_{cit}^p \geq N_{cit}^{p+1} \), it is given by

   \[
   h_A = \max \{ p \mid p \leq N_{cit}^p \}. \tag{2}
   \]

   This is proportional to \( \sqrt{N_{cit}^A} \), times a factor that penalises authors that write a small number of highly cited papers [4].

As we will see, the average number of authors per paper and of references per paper increased, in the last 20 years, by one to a few per-cent a year, and is significantly different in different communities. Following basic common-sense, we propose an improved metric that renormalises away such factors, and that cannot be artificially inflated adding more references and/or more co-authors.

4. **Number of individual citations**

   A citation from paper \( p' \) to paper \( p \) is weighted as the inverse of the number of references

---

¹For other notable digital libraries and databases of research in various fields see ref.s [10–23].
²The most relevant literature before ≈ 1970 has been added and is still being added on a request base to INSPIRE.
$N_{p}^{ref}$ of paper $p'$. Furthermore, the citations received by a paper $p$ are equally shared among its $N_{p}^{aut}$ authors.\(^3\) In formulæ:

$$
N^{icit}_p = \sum_{p' \rightarrow p} \frac{1}{N_{p'}^{ref}}, \quad N^{icit}_A = \sum_{p \in A} \frac{N^{icit}_p}{N_{p}^{aut}}.
$$

(3)

All the above metrics are defined “locally”, i.e. they can be computed for a paper/author without knowing anything about all other papers but the ones that cite it. Therefore, they all potentially suffer from the problem that even small sub-communities can inflate their indicators. To overcome this problem one needs to define global measures, i.e. measures that know about the whole community. The simplest such global measure of impact is given by the PageRank algorithm, introduced in 1996 by Larry Page and Sergey Brin, the funders of Google \([24, 25]\).\(^4\) For a pedagogical introduction to the PageRank see ref.s \([26,27]\). Applications of the PageRank algorithm to citations network have already been considered, for instance, in ref.s \([28–31]\). More advanced ranking algorithms, based on integrated bibliographic information have also been proposed in ref.s \([32–34]\).

5. PaperRank

The PaperRank $R_p$ of paper $p$ and the PaperRank $R_A$ of author $A$ are defined as

$$
R_p = \sum_{p' \rightarrow p} \frac{R_{p'}}{N^{ref}_{p'}}, \quad R_A = \sum_{p \in A} \frac{R_{p}}{N_{p}^{aut}}.
$$

(4)

Namely, citations from papers $p'$ are weighted proportionally to their ranks $R_{p'}$, that get thereby determined trough a system of linear equations. The PaperRank provides a metric which cannot be easily artificially inflated, because it is the bibliometric estimator of a physical quantity: how many times each paper is read.

As we will see, the PaperRank singles out notable old papers which often do not have many citations. However, given that citations are time-ordered (newer papers cite older ones), the rank reduces to a weighted sum over citation descendants (a combination of “citations of citations”) which needs about 10-20 years to become a better indicator than the number of individual citations. In order to use information from the past, we define an alternative AuthorRank based on citations among authors.

---

\(^3\)We assume that authors contributed equally because in fundamental physics authors are usually listed alphabetically, with no information about who contributed more. The factor $1/N_{p}^{aut}$ is then dictated by conservation laws.

\(^4\)Even if the word “Page” in PageRank may seem to refer to webpages, the name of the algorithm originates from the name of one of its inventors, Larry Page.
6. **AuthorRank**

We define the citation matrix which counts all individual citations from author $A'$ to $A$

$$N_{A'\rightarrow A}^{\text{cit}} = \sum_{p_{A'} \rightarrow p_A} \frac{1}{N_{p_A}^{\text{aut}} N_{p_{A'}}^{\text{aut}}} \frac{1}{N_{p_{A'}}^{\text{ref}}} ,$$

where the sum runs over all papers $p_{A'}$ of author $A'$ that cite papers $p_A$ of $A$. We then define the AuthorRank $R_A$ as

$$R_A = \sum_{A'} R_{A'} C_{A'\rightarrow A} , \quad C_{A'\rightarrow A} = \frac{N_{A'\rightarrow A}^{\text{cit}}}{\sum_{A''} N_{A'\rightarrow A''}^{\text{cit}}} ,$$

namely, as the principal eigenvector of the right stochastic matrix $C_{A'\rightarrow A}$, which (thanks to the normalization of each row provided by the sum in the denominator) tells the percentage of individual citations to $A$ among all individual citations of $A'$. The AuthorRank gives more weight to citations coming from highly cited authors. We also use the AuthorRank of authors to define an improved ranking of papers, that we call AuthorRank of papers, as

$$R_p = \sum_{p' \rightarrow p} \sum_{A' \in p'} \frac{R_A}{N_{p'}^{\text{aut}} N_{p'}^{\text{ref}}} .$$

All above metrics can optionally be modified by removing self-citations. Furthermore, one can similarly remove “citation cartels” among 3 authors, among 4 authors, etc. The mathematical problem of removing all circular citations has a simple solution: money. You don’t get richer by giving money to yourself or by circulating money with friends. This leads us to the definition of an additional metric:

7. **Citation-coin**

Author $A$ ‘owes’ the number $\Phi_A$ of individual citations received minus the number of individual citations given:

$$\Phi_A = \sum_{A'} (N_{A'\rightarrow A}^{\text{cit}} - N_{A\rightarrow A'}^{\text{cit}}) = N_A^{\text{cit}} - \sum_{p \in A} \frac{1}{N_{p}^{\text{aut}}} .$$

This metric penalises authors who write many small papers which receive few citations from others.

An important property of all above metrics is that they can be computed in practice. Our metrics are intensive: so they can be used to evaluate groups, such as journals, institutes, countries, etc. by simply summing over their members. Furthermore they can be restricted to any specific time period, e.g. after year 2000.

The paper is structured as follows. In section 2 we describe the INSpire and arXiv databases and their main features and trends. In section 3 we introduce the PaperRank $R$, discuss its
The number of papers in each category is computed considering only the main category and ignoring cross-list, so that no paper can belong to more than one gran category.

features and properties, and rank all papers in INSpire. In section 4 we introduce the number of individual citations $N_{\text{cit}}$, the AuthorRank $\mathcal{R}$ and the citation coin $\mathcal{C}$, discussing their features and properties, and rank all authors in INSpire. In section 5 we apply these measures to rank groups: institutions, towns, countries, continents, journals, genders. Conclusions are presented in section 6. Technical details are presented in appendix A.

Several of our results with complete tables are available at the webpage [35].

2 The INSpire and arXiv databases

The open-source INSpire bibliographic database [5] covers fundamental physics world-wide. INSpire presently contains about $10^6$ papers, $3 \cdot 10^7$ references, $10^5$ authors, $10^4$ institutions, and $4 \cdot 10^3$ journals. INSpire started around 1965, but it also contains some notable older papers. INSpire maps papers, authors, institutes (affiliations), and journals to record IDs (integer numbers) thereby addressing the problem of name disambiguation [36].

Starting from 1995, preprints for most of the papers contained in INSpire are available through arXiv.org [12], which also covers fields beyond fundamental physics, so that not all of the arXiv database is included into the INSpire one. The arXiv also provides a classification in terms of categories, some of which contain sub-classes. The arXiv categories and the number
Figure 2: Number of papers per year (green), average number of references (red), of authors (magenta), of citations (blue), of citations among published papers (blue dashed). Fig. 3 shows the same trends within arXiv categories.

of papers in each of them are shown in the left histogram in figure 3. The right histogram shows the fraction of papers in the various categories included in InSpire. We also often show results for the main arXiv categories inside InSpire, defined as the arXiv categories with more than $10^3$ papers, and with a fraction included in InSpire larger than 50%: hep-ex (high-energy experiment), hep-ph (high energy theory/phenomenology), hep-th (high energy theory), astroph (astrophysics and cosmology), hep-lat (lattice field theory), nucl-ex (nuclear experiment), nucl-th (nuclear theory), gr-qc (general relativity and quantum cosmology). Details of the dataset we consider and technical issues about the InSpire database are discussed in appendix A.

2.1 Main trends in the fundamental physics literature

Figure 2 shows the time evolution of some main factors: number of papers per year (which increased by 5%/yr); average number of references per paper (increased by 3%/yr); number of citations per paper (roughly constant, taking into account that recent papers, published in the past $\approx 15$ years, necessarily received less citations); number of authors per paper (increased from few to tens). We also see that most citations go to published papers.

Fig. 3 shows the same trends within the main arXiv representative categories, showing that papers with an increasingly large number of authors lie in experimental categories (hep-ex, nucl-ex, astro-ph), while papers in other fields keep having, on average, 2 – 3 authors.

Fig. 4 shows the ‘birth’ and ‘death’ rates within the main arXiv categories as function of time: in green the percentage of authors who published in year $y$ but not in year $y - 1$; in red

\[^5\]The full list of sub-classes can be found in ref. [37].
the percentage of authors who published in year \( y - 1 \) but not in year \( y \). The balance is stable, with a significant growth of hep-ex when LHC started, and of astro-ph until 2010.

The Gini coefficient is used in economy as a measure of inequality of wealth. Typical occidental countries have Gini coefficients between 0.2 and 0.4. The typical Gini coefficient of citations is 0.7. This means that few papers get a lot of citations: 4% of papers have more than 100 citations, and they receive half of the total citations; half of the papers have less than 4 citations, and they receive 2% of the total citations (see also ref. [38]).

3 Ranking papers

3.1 PaperRank

Given a citation network of \( N_{pap} \) papers, the PaperRank \( R_p \) of each paper \( p \) is defined by

\[
R_p = \varphi \sum_{p' \rightarrow p} \frac{R_{p'}}{N_{p'}} + \alpha (1 - \varphi),
\]  

(9)
Figure 4: Percentage of authors appeared or disappeared each year within the arXiv categories fully covered by InSpire.

where \( N_{\text{ref}}' \) is the total number of references of each paper \( p' \) that cites \( p \). Equation (9) is linear, so its solution is unique, and can be computed efficiently iteratively [27].

We elaborate on its meaning. Eq. (9) contains two arbitrary constants \( \alpha \) and \( \wp \). The constant \( \alpha \) just fixes the overall normalization \( \sum_p R_p = R_{\text{tot}} \). When applied to internet, \( R_p \) describes the probability that site \( p \) is visited and it is convenient to normalize it to one. We choose \( R_{\text{tot}} \) equal to the total number of citations \( R_{\text{tot}} = \sum_p N_{\text{cit}}' \), in order to allow for an easier comparison between the number of citations received by a paper and its PaperRank \( R_p \). With this normalisation \( R_p \) grows with time, as newer papers appear.

Viewing the rank as the probability that a paper is read, the parameter \( \wp \) splits it into two contributions: the first term is the probability that a reader reaches a paper by following a reference to it; the second term, equal for all papers, simulates readers that randomly browse the literature.

- In the limit \( \wp = 0 \) the first contribution vanishes, and all papers have a common rank. At first order in small \( \wp \ll 1 \), \( R_p \) starts discriminating the papers \( p \):

\[
R_p(\wp) \approx R_{\text{tot}} \frac{1 + \wp N_{\text{cit}}'}{N_{\text{pap}}} 1 + \wp, \quad (10)
\]

where \( N_{\text{cit}}' \) is the number of individual citations received by \( p \) defined in eq. (3), which obeys \( \sum_p N_{\text{cit}}' = N_{\text{pap}} \).
• In the limit $\varphi = 1$ the second contribution in eq. (9) vanishes, and $R_p$ only depends on the structure of the network, provided that no closed sub-networks and dead-ends exist [27]. Recursive computations of $R_p$ become slower as $\varphi \to 1$.

Data about downloads of scientific articles would allow to extract the value of $\varphi$ that better fits the observed reading rate; however such data are not available.\(^6\) We use a large $\varphi = 0.99$, such that the first contribution in eq. (9) dominates for all relevant authors.

### 3.2 PaperRank of papers: results

We compute the PaperRank by constructing a graph (and its transition matrix) having all papers as nodes and citations as links. We consider the full InSpire database, as detailed in appendix A. Generally, a few hundred iterations of eq. (9) are necessary for a percent level convergence. We also check our result against the PageRankCentrality Mathematica function finding agreement. The computation takes a few minutes on a laptop computer.

Table 1 (table 2) shows the top-cited (top-ranked) papers in the InSpire database. Top-ranked papers correspond to the papers with top PaperRank and tend to be old famous ones, even with a relatively small number of citations. Top-cited papers, ranked with the usual counting of the number of citations, tend to be modern, in the view of the inflation in the rate of citations. The same effect was observed in ref. [28], which applied the PageRank algorithm to the sub-set of papers published on Physical Review.

The difference between the two rankings is partly due to the fact that PaperRank penalises recent papers. Papers tend to accumulate citations for about 10-20 years, while the rank continues growing with time, and is highly suppressed for younger papers.

This also means that the PaperRank defined in eq. (9) needs 10-20 years before providing a better metrics than the number of citations. This is proven in section 3.3 where we show that, for a time-ordered network (such as the network of citations), the PaperRank reduces to the number of citations-of-citations.

### 3.3 PaperRank as the number of citations-of-citations

Internet allows for reciprocal links among pages, and the PageRank captures in a simple way the self-interacting system. Citations among scientific papers are instead time-ordered, forming an acyclic network. In the limit where citations of older papers to newer papers are ignored,\(^7\) no loops are possible within the network, and the implicit definition of the rank $R_p$ of eq. (9)

\(^6\)arXiv.org does not make public the number of downloads, to avoid that some authors might be tempted to artificially enhance the downloads of their articles.\(^7\)We enforced time-ordering within the citations, deleting from the InSpire database a small number of ‘acausal’ citations, where older papers cite newer papers (see appendix A for details). Since older papers tend to accumulate large ranks, acausal citations can artificially inflate the rank of a few recent papers.
1. The Large $N$ limit of superconformal field theory (Maldacena, 1998)
2. A Model of Leptons (Weinberg, 1968)
3. Measurements of Omega and Lambda from 42 (Perlmutter, 1999)
4. Observational evidence from supernovae (Riess, 1998)
5. CP Violation in the Renormalizable Theory (Kobayashi, 1973)
6. PYTHIA 6.4 Physics and Manual (Sjostrand, 2006)
7. GEANT4: A Simulation toolkit (Amako, 2003)
8. Anti-de Sitter space and holography (Witten, 1998)
9. First year Wilkinson Microwave Anisotropy (Spergel, 2003)
10. A Large mass hierarchy from a small extra dimension (Randall, 1999)

Table 1: Top-cited (highest number of citations) papers in the InSpire database.

| Title                                      | 1st author | $N_{aut}$ | date | $N_{cit}$ | $R_p$ | $\mathcal{R}_p$ |
|--------------------------------------------|------------|-----------|------|----------|-------|-----------------|
| The Large $N$ limit of superconformal field | J.M. Maldacena | 1 | 1998 | 13242 | 5239 | 17779 |
| A Model of Leptons                         | S. Weinberg | 1 | 1968 | 11157 | 38925 | 34449 |
| Measurements of Omega and Lambda from 42  | S. Perlmutter | 32 | 1999 | 10485 | 3251 | 5908 |
| Observational evidence from supernovae f  | A. G. Riess | 20 | 1998 | 10306 | 3359 | 5779 |
| CP Violation in the Renormalizable Theory  | M. Kobayashi | 2 | 1973 | 9779 | 9778 | 17611 |
| PYTHIA 6.4 Physics and Manual              | T. Sjostrand | 3 | 2006 | 9648 | 2561 | 3024 |
| GEANT4: A Simulation toolkit               | K. Amako | 125 | 2003 | 8979 | 4192 | 1529 |
| Anti-de Sitter space and holography        | E. Witten | 1 | 1998 | 8659 | 3510 | 11397 |
| First year Wilkinson Microwave Anisotropy  | D. N. Spergel | 17 | 2003 | 8277 | 2427 | 5283 |
| A Large mass hierarchy from a small extr   | Lisa. Randall | 2 | 1999 | 7881 | 2667 | 5853 |

Table 2: Top-ranked (highest PaperRank) papers in the InSpire database.

| Title                                      | 1st author | $N_{aut}$ | date | $N_{cit}$ | $R_p$ | $\mathcal{R}_p$ |
|--------------------------------------------|------------|-----------|------|----------|-------|-----------------|
| A Model of Leptons                         | S. Weinberg | 1 | 1968 | 11157 | 38925 | 34449 |
| Particle Creation by Black Holes           | S. W. Hawking | 1 | 1974 | 6833 | 7368 | 23268 |
| Unity of All Elementary Particle Forces    | H. M. Georgi | 2 | 1974 | 4517 | 9023 | 22309 |
| A Planar Diagram Theory for Strong Inter   | G. 't Hooft | 1 | 1974 | 4400 | 7061 | 22019 |
| Confinement of Quarks                      | K. G. Wilson | 1 | 1974 | 4580 | 20257 | 20138 |
| Weak Interactions with Lepton-Hadron Sym   | S. L. Glashow | 3 | 1970 | 5623 | 16914 | 19709 |
| Pseudoparticle Solutions of the Yang-Mil   | A. A. Belavin | 4 | 1975 | 2675 | 11316 | 19462 |
| Symmetry Breaking Through Bell-Jackiw An   | G. 't Hooft | 1 | 1976 | 3389 | 6887 | 18032 |
| The Large $N$ limit of superconformal field | J. M. Maldacena | 1 | 1998 | 13242 | 5239 | 17779 |
| CP Violation in the Renormalizable Theory  | M. Kobayashi | 2 | 1973 | 9779 | 9778 | 17611 |

Table 3: Top-referred (highest AuthorRank) papers in the InSpire database.

11
Figure 5: Contributions of each generation to the citation chain of some notable papers.

can be converted into the following explicit expression \(^8\)

\[
R_p \propto \sum_{g=0}^{\infty} \varphi^g \sum_{p_g \rightarrow \cdots \rightarrow p} \frac{1}{N_{\text{ref}}^{p_g}} \cdots \frac{1}{N_{\text{ref}}^{p_1}}.
\]

(11)

Basically, \(R_p\) counts the number of citations-of-citations up to generation \(g\). In the above expression, the term with

- \(g = 0\) contributes as unity, and accounts for the constant term in eq. (9), which is negligible for papers that receive citations from others;

- \(g = 1\) contributes with the number of individual citations \(N_{\text{p}}^\text{cit}\) as in eq. (10): the sum runs over ‘first generation’ papers \(p_1\) that cite the paper \(p\);

- \(g = 2\) corresponds to ‘second generation’ papers \(p_2\) that cite the papers \(p_1\) that cite \(p\);

- \(g = 3\) corresponds to ‘third generation’ papers \(p_3\) that cite the papers \(p_2\) that cite the papers \(p_1\) that cite \(p\);

- for generic \(g\) the sum runs over all papers \(p_g\) that cite paper \(p\) in \(g\) steps.

A paper \(q\) can appear multiple times in different generations \(g\), corresponding to all possible citations paths from \(q\) to \(p\). Eq. (11) shows that \(\varphi < 1\) gives a cut-off on the number of generations that one wants to consider, and that \(R_p(\varphi)\) grows with \(\varphi\), and with time.

- At one extremum, \(\varphi \to 0\), the rank \(R_p\) reduces to the “number of children” \(N_{\text{p}}^\text{cit}\), without checking if they are successful. Papers on hot topics can fastly accumulate many citations, even if later the hot topic becomes a dead topic. Too recent papers are penalised.

\(^8\)This can be proofed by substituting eq. (11) into eq. (9). A physicist can view in eq. (11) a path-integral within the network.
Table 4: The top-cited, top-ranked, top-referred paper written in each decennium among those listed in InSpire (which is highly incomplete before 1960).

- At the other extremum, $\varphi \to 1$, the rank $R_p$ becomes the Adamo number: it counts descendants. Seminal papers that open new successful fields are rated highly, and their rank continues to grow. Too recent papers are highly penalised.

The PaperRank in eq. (11) splits papers into two qualitatively different categories:

- Sub-critical papers that get some attention and get forgotten: this happens when, after a long time (tens of years), the sum over $g$ remains dominated by the first generation.

- Super-critical papers, that make history. If the citation rate is high enough, it can sustain a ‘chain reaction’, such that late generations keep contributing significantly to the sum over generations. At the same time, the original paper gets summarized in books and ceases to be cited directly.

Fig. 5 shows, for a few notable papers how much different generations $g$ contribute to the sum in eq. (11). We see that about 10 generations contribute significantly for old top-ranked papers, while the 1st generation provides the dominant contribution to recent top-cited papers.\(^9\)

\(^9\)We computed $R_p(\varphi)$ analytically as function of $\varphi$ for all papers $p$ using eq. (11) with the following ‘pruning’ algorithm. To start, one finds all papers with no citations and eliminates them from the database, after assigning
Table 5: Top-referred papers written after 2010 with less than 10 authors.

| Title                                                                                  | 1st author | N$_{aut}$ | date | N$_{cit}$ | R$_{p}$ | R$_{p}$ |
|----------------------------------------------------------------------------------------|------------|-----------|------|-----------|--------|--------|
| Black Holes: Complementarity or Firewall                                                | A.Almheiri | 4         | 2012 | 655       | 121    | 2668   |
| Topological Insulators                                                                  | C.L.Kane   | 2         | 2010 | 1141      | 430    | 673    |
| On the Origin of Gravity and the Laws of                                                | E.P.Verlinde | 1        | 2010 | 642       | 139    | 637    |
| Topological insulators and superconductors                                              | X.L.Qi     | 2         | 2011 | 895       | 243    | 630    |
| MadGraph 5: Going Beyond                                                                | J.Alwall   | 5         | 2011 | 2641      | 407    | 625    |
| Shapiro Delay Measurement of A Two Solar                                               | S.M.Ransom | 5         | 2011 | 1559      | 240    | 580    |
| Towards an understanding of jet substructure                                           | M.Dasgupta | 4         | 2013 | 150       | 19     | 578    |
| New parton distributions for collider ph                                                | H.L.Lai    | 7         | 2010 | 2370      | 388    | 547    |
| Investigating the near-criticality of the                                               | D.Buttazzo | 7         | 2014 | 701       | 83     | 544    |
| An Apologia for Firewalls                                                               | A.Almheiri | 5         | 2013 | 232       | 41     | 500    |

3.4 Top-referred (recent) papers

As explained in the previous section, the PaperRank can single out some notable papers with few citations, provided that they are old. However, when applied to recent papers (less than 10-20 years old), the PaperRank becomes highly correlated to the number of (individual) citations, and therefore cannot perform better.

We thereby propose an early-alert indicator that recovers some information from the past. First, we compute a rank among authors using all-time data. In particular, we adopt the AuthorRank $R_A$ anticipated in eq. (6) and better discussed in the next section. Next, we use such rank to weight citations to papers, as in eq. (7). This means that we give more weight to citations from authors with higher $R_A$, implementing a sort of representative democracy. We dub papers with top AuthorRank of papers $R_p$ as ‘top-referred papers’. Table 3 shows the all-time list.

Table 4 shows the top-cited, top-ranked and top-referred papers published within each decennium, based on all subsequent citations. For recent papers, top-cited and top-ranked tend to be dominated by manuals of useful computer codes and by reviews.

We finally use $R_p$ to find top-referred recent papers: table 5 shows the top-referred papers published after 2010 and with less than 10 authors (because our goal is to identify notable recent papers less known than discoveries made by big experimental collaborations). Furthermore, we here removed self-citations, to avoid the list to be dominated by notable authors citing themselves.

3.5 Paper metrics: correlations

The left panel of figure 6 shows the correlations among the traditional counting of citations $N_{cit}$, the number of individual citations $N_{cit}^p$, the PaperRank $R_p$ and the AuthorRank of papers $R_p$, in the whole InSpire database. The number of individual citations is highly correlated with the number of citations. Indeed, for papers, individual citations are just citations divided their contributions to the $R_p$ of their references. The process is iterated. About 1000 iterations are needed to prune all the InSpire citation tree to nothing, obtaining $R_p(\varphi)$ as a power series in $\varphi$. 

14
Figure 6: Left(right) table: correlations between indices of papers(authors). For papers the correlation is considered separately for the whole InSpire and for the eight main arXiv categories.

Table 6: Authors listed according to traditional biblio-metric indices: total number of papers (left column), of citations (middle), h-index (right).

by the number of references of the citing papers so that uncorrelation is proportional to the variance around the average of the number of references per paper. The PaperRank and the AuthorRank are less correlated to the number of (individual) citations and to each other and represent fairly independent indices for ranking papers.

4 Ranking authors

We start from the simplest and most naive metrics: in the left column of table 6 we list the authors with most papers. Within the InSpire database, they are all experimentalists that participate in large collaborations with many co-authors. The extreme case are the ATLAS and CMS collaborations with \( \sim 10^3 \) papers and \( \sim 10^3 \) authors.

In the middle column of table 6 we show the top-cited authors: again they are experimentalists that participate in large collaborations. The citations of author \( A \) are counted in the usual way: summing the citations received by all papers that include \( A \) as author, as in eq. (1). The right column of table 6 shows the \( h \) index.
Table 7: Authors sorted according to their number of individual citations \( N_{A_{\text{cit}}} \) for the whole InSpire database (left), after year 2000 (middle), and after year 2010 (right).
Clearly, the total number of papers and of citations and the \( h \) index ceased to be relevant for experimentalists in view of the large number of co-authors. This shows the need for an improved metrics that corrects for the inflation in the number of co-authors, since attributing the whole paper to each co-author is contrary to common-sense.\(^{10}\)

### 4.1 Sharing among co-authors

A correct metrics is obtained by attributing a fraction \( p_A \) of any given paper to each author \( A \), and imposing the sum rule \( \sum p_A = 1 \). The fractions \( p_A \) should tell how much each author contributed to the paper. In the absence of this information, we assume that each co-author contributed equally, so that \( p_A = 1/N_{\text{aut}} \).\(^{11}\)

Taking into account this factor, the total number of ‘individual papers’ of author \( A \) is given by \( \sum_{p \in A} 1/N_{\text{aut}} \). The same sharing among co-authors is applied to citations. The number of ‘individual citations’ received by author \( A \) is defined by summing over all its papers \( p \) taking into account that citations are shared among co-authors, and weighted inversely to the number of references:

\[
N_{\text{icit}}^A \equiv \sum_{p \in A} \frac{1}{N_{\text{aut}}^p} \left( \frac{1}{\sum_{p' \rightarrow p} (\text{Number of references of } p')} \right).
\]

(12)

In the same way, we share the rank \( R_p \) of each paper equally among its authors. The rank \( R_p \) of a paper approximates a physical quantity: how many times the paper is read. The rank \( R_A \) of an author inherits the same meaning: it tells the visibility of any author \( A \), obtained by summing the visibility of its papers \( p \) as in eq. (4), i.e.\(^{12}\)

\[
R_A = \text{(PaperRank of author } A) = \sum_{p \in A} \frac{(\text{PaperRank of paper } p)}{(\text{Number of authors of paper } p)}.
\]

(13)

As discussed in section 3.1 we consider \( \varphi = 0.99 \).

### 4.2 PaperRank of authors: results

We now apply the improved metrics described in the previous section to the INSPIRE database. The left column of table 7 lists the authors with the highest number of individual citations. Two factors differentiate citations from individual citations. First, dividing by the number of references (first factor in the denominator in eq. (12)) counter-acts the inflation in the total number of citations. This factors mildly penalises authors working in sectors (such as hep-ph) where papers have a larger average number of references. Second, dividing by the number

---

\(^{10}\) No algorithm can discriminate among authors that only write papers in a large collaboration. Sub-groups of some collaborations write ‘internal notes’, which however remain private within the collaboration.

\(^{11}\) Weighting authors proportionally to their AuthorRank gives a non-linear system of equations, with singular solutions where a few notable authors collect all the weight of large collaborations.

\(^{12}\) For different methods see ref.s [29, 31].
Table 8: Authors sorted according to the PaperRank of authors, $R_A = \sum_{papers} R_p / N_{aut}$ for the whole InSpire database (left), after year 2000 (middle), and after year 2010 (right).

| InSpire name | All InSpire | After 2000 | After 2010 |
|--------------|-------------|------------|------------|
| 1 S.Weinberg.1 | 201372 | 2560 | 629.6 |
| 2 J.S.Schwinger.1 | 172155 | 2371 | 517.1 |
| 3 R.P.Feynman.1 | 127049 | 2242 | 477.4 |
| 4 M.Gell.Mann.1 | 118127 | 2228 | 446.7 |
| 5 P.A.M.Dirac.1 | 93587 | 2172 | 444.9 |
| 6 E.Witten.1 | 86077 | 2130 | 437.7 |
| 7 G.Hooft.1 | 84539 | 2070 | 426.4 |
| 8 Abdus.Salam.1 | 79862 | 2069 | 419.7 |
| 9 C.N.Yang.1 | 79224 | 2020 | 409.4 |
| 10 H.A.Bethe.1 | 78725 | 1992 | 404.6 |
| 11 S.L.Adler.1 | 65420 | 1924 | 398.9 |
| 12 T.D.Lee.1 | 62141 | 1902 | 395.8 |
| 13 Yoichiro.Nambu.1 | 61056 | 1866 | 391.3 |
| 14 E.P.Wigner.1 | 60278 | 1862 | 383.4 |
| 15 A.M.Polyakov.1 | 56791 | 1841 | 382.7 |
| 16 W.Heisenberg.1 | 56487 | 1806 | 375.5 |
| 17 S.L.Glashow.1 | 55695 | 1752 | 372.5 |
| 18 K.G.Wilson.1 | 53351 | 1732 | 359.1 |
| 19 J.D.Bjorken.1 | 54473 | 1702 | 351.5 |
| 20 Max.Born.1 | 51519 | 1680 | 345.7 |
| 21 S.W.Hawking.1 | 50582 | 1604 | 345.7 |
| 22 B.Zumino.1 | 50519 | 1602 | 343.6 |
| 23 S.Mandelstam.1 | 48738 | 1599 | 341.4 |
| 24 F.J.Dyson.1 | 48250 | 1579 | 335.1 |
| 25 P.W.Higgs.1 | 43760 | 1553 | 328.8 |
| 26 D.J.Gross.1 | 43454 | 1546 | 322.8 |
| 27 J.H.Schwarz.1 | 42715 | 1511 | 312.7 |
| 28 J.A.Wheeler.1 | 42398 | 1449 | 320.4 |
| 29 R.W.Jackiw.1 | 41256 | 1440 | 318.7 |
| 30 G.F.Chew.1 | 40569 | 1434 | 310.8 |

The left column of table 8 shows the top-ranked authors in the InSpire database. The rank identifies some older notable authors who have less citations than modern authors, given the increase in the rate of papers and of citations.

Anyhow, the main interest of our study is not re-discovering Feynman. We want to see if our metrics do a better job than just citation counts in identifying modern authors with a high impact. To achieve this, we set a lower cut-off on the publication year. We restrict the list to papers published ‘After 2000’ (middle column of table 8) and ‘After 2010’ (right column of table 8).

While switching from citations to individual citations is an obvious improvement, we find that the rank does not improve over individual citations (they are strongly correlated) when restricting to recent papers. As already discussed for papers, about 10-20 years are needed...
Table 9: Authors sorted according to their author rank $R_A$.

| InSpires name | All | InSpires name | After 2000 | InSpires name | After 2010 |
|---------------|-----|---------------|------------|---------------|------------|
| P.A.M.Dirac.1  | 239145 | T.Sjostrand.1  | 13918 | N.Kidonakis.1  | 5685 |
| E.Witten.1     | 147226 | E.Witten.1     | 13699 | A.A.Abdo.1     | 3093 |
| S.Weinberg.1    | 131054 | V.Springel.1   | 11771 | P.Z.Slans.1    | 2998 |
| G.tHooft.1      | 98505  | P.Z.Slans.1    | 11245 | M.Czakon.1     | 2671 |
| Albert.Einstein.1 | 91407 | J.M.Maldacena.1 | 10871 | M.Luscher.1    | 2522 |
| J.S.Schwinger.1 | 81662  | S.Mrenna.1     | 10041 | E.Witten.1     | 2349 |
| S.W.Hawking.1   | 69961  | C.Vafa.1       | 9283  | X.L.Qi.1       | 2184 |
| A.M.Polyakov.1  | 63934  | G.P.Salam.1    | 8837  | G.Harry.1      | 2133 |
| R.P.Feynman.1   | 61184  | P.Nason.1      | 8490  | P.Nason.1      | 2033 |
| M.Gell.Mann.1   | 57174  | D.T.Son.1      | 8083  | C.L.Kane.1     | 1988 |
| Max.Born.1      | 55314  | B.R.Webber.1   | 7657  | A.Mitov.1      | 1918 |
| C.N.Yang.1      | 54547  | A.D.Martin.1   | 7569  | J.Rojo.1       | 1898 |
| K.G.Wilson.1    | 50032  | J.A.M.Vermaseren.1 | 7270 | S.Sachdev.1    | 1810 |
| H.A.Bethe.1     | 46454  | S.D.M.White.1  | 6933  | F.Maltoni.1    | 1788 |
| Enrico.Fermi.1  | 45455  | S.Frixione.1   | 6845  | G.P.Salam.1    | 1732 |
| T.D.Lee.1       | 42137  | N.Berkovits.1  | 6766  | C.de.Rham.1    | 1716 |
| L.Susskind.1    | 41491  | L.E.Hernquist.1| 6762  | J.M.Maldacena.1| 1712 |
| S.L.Glashow.1   | 40930  | M.Cacciari.1   | 6673  | K.Somiya.1     | 1647 |
| Abdus.Salam.1   | 39924  | A.Shoke.Sen.1  | 6625  | Alexander.Romanenko.1 | 1621 |
| R.W.Jackiw.1    | 38934  | A.Loeb.1       | 6585  | P.Huber.2      | 1617 |
| F.A.Wilczek.1   | 38674  | N.Kidonakis.1  | 6561  | H.T.Janka.1    | 1613 |
| H.M.Georgi.1    | 36890  | Wayne.Hu.1     | 6295  | S.Forte.2      | 1611 |
| S.L.Adler.1     | 35532  | J.Polchinski.1 | 6173  | John.M.Campbell.1 | 1606 |
| D.J.Gross.1     | 35278  | A.C.Fabian.1   | 6118  | Emanuele.Re.1  | 1587 |
| S.R.Coleman.1   | 34428  | F.Aharonion.1  | 6056  | R.K.Ellis.1    | 1546 |
| A.D.Linde.1     | 34189  | B.T.Draine.1   | 6036  | S.O.Moch.1     | 1514 |
| J.D.Bjorken.1   | 33920  | F.Zimmermann.1 | 6036  | Xiao.Gang.Wen.1| 1499 |
| T.Sjostrand.1   | 31932  | R.S.Thorne.1   | 5947  | E.Gross.2      | 1487 |
| W.Heisenberg.1  | 31057  | M.Luscher.1    | 5892  | O.Vitella.1    | 1484 |
| S.Mandelstam.1  | 30984  | P.Kroupa.1     | 5879  | A.Vishwanath.1 | 1449 |

4.3 Author Rank

As outlined in the introduction, the citation matrix between authors (properly normalized) $C_{A'\to A}$ defined in eq. (6) allows to define an AuthorRank as

$$R_A = \varphi \sum_{A'} R_{A'} C_{A'\to A} + \alpha (1 - \varphi) \, .$$

(14)

where the second term gives a constant weight to each author, independently from the number and quality of its papers. While formally analogous to the ranking of papers, this ranking
of authors is not a model of a physical process, because one reads papers, not authors. The
graph corresponding to the matrix \( C_{A' \rightarrow A} \) contains cycles and also loops on the same note
(self-citations): here we use \( \wp = 0.9 \), such that self-citations cannot boost the AuthorRank by
more than one order of magnitude.

The left column of table 9 shows the all-time AuthorRank. We see the emergence of old
authors such as Einstein, which were absent from previous top-rankings because poorly covered
and cited in the too recent InSpire database. Of course, the incompleteness of InSpire before
\( \sim 1960 \) makes results about older authors semi-quantitative.

This issue is avoided in the other columns of table 9, where we show the AuthorRank
recomputed by restricting to recent papers only.

### 4.4 Removing self-citations and cartels: the Citation Coin

One of the unsatisfactory aspects of previous metrics is the effect of self-citations.

On short time-scales the PaperRank \( R \) and the number of (individual) citations are strongly
correlated, so they can be similarly inflated through self-citations. Only on longer time-scales \( R \)

| InSpire name | All | InSpire name | After 2000 | InSpire name | After 2010 |
|--------------|-----|--------------|------------|--------------|------------|
| 1 E.Witten. | 3085 | T.Sjostrand. | 183.8 | C.de.Rham. | 23.9 |
| 2 S.Weinberg. | 2072 | S.Mrenna. | 174.3 | M.Luscher. | 22.9 |
| 3 G.tHooft. | 1237 | P.Z.Skands. | 162.7 | N.Kidonakis. | 21.9 |
| 4 S.W.Hawking. | 1050 | J.M.Maldacena. | 160.5 | E.P.Verlinde. | 19.8 |
| 5 A.M.Polyakov. | 866.0 | G.P.Salam. | 118.3 | A.A.Abdo. | 19.6 |
| 6 J.Schwinger. | 674.3 | E.Witten. | 114.1 | P.Z.Skands. | 18.3 |
| 7 J.M.Maldacena. | 653.3 | S.D.Odintsov. | 110.0 | T.J.Stelzer. | 16.1 |
| 8 T.Sjostrand. | 635.0 | D.T.Son. | 108.4 | J.M.Maldacena. | 16.0 |
| 9 R.W.Jackiw. | 590.0 | V.Springel. | 107.5 | S.F.Hassan. | 16.0 |
| 10 S.R.Coleman. | 578.1 | M.Cacciari. | 105.3 | G.Harry. | 15.5 |
| 11 S.L.Glashow. | 578.3 | S.Nojiri. | 104.3 | M.Czakon. | 15.2 |
| 12 F.A.Wilczek. | 573.3 | P.Nason. | 98.5 | F.Maltoni. | 14.9 |
| 13 H.M.Georgi. | 571.1 | T.Padmanabhan. | 93.8 | J.Alwall. | 14.8 |
| 14 P.A.M.Dirac. | 565.7 | N.Arkan.Hamed. | 88.2 | G.P.Salam. | 14.7 |
| 15 L.Susskind. | 561.9 | C.Vafa. | 84.8 | R.A.Rosen. | 14.1 |
| 16 A.D.Linde. | 548.8 | V.A.Kostelecky. | 76.4 | G.Gabadadze. | 14.0 |
| 17 N.Seiberg. | 547.0 | Ashoke.Sen. | 75.8 | A.Mitov. | 13.8 |
| 18 K.G.Wilson. | 509.4 | P.Horava. | 74.3 | M.Cacciari. | 12.9 |
| 19 D.J.Gross. | 501.9 | J.Polchinski. | 72.5 | O.Mattelaer. | 12.5 |
| 20 M.Luscher. | 446.9 | B.R.Webber. | 67.8 | P.Nason. | 12.4 |
| 21 C.Vafa. | 431.9 | A.Strominger. | 67.1 | S.S.Meyer. | 11.6 |
| 22 R.P.Feynman. | 423.9 | M.Luscher. | 67.0 | M.R.Nolta. | 11.6 |
| 23 R.L.Jaffe. | 405.8 | S.S.Gubser. | 64.8 | M.Maltoni. | 11.5 |
| 24 B.Zumino. | 390.1 | S.Frixione. | 64.6 | N.Jarosik. | 11.5 |
| 25 J.D.Bjorken. | 383.8 | B.Ratra. | 64.4 | C.L.Kane. | 11.3 |
| 26 M.Gell.Mann. | 374.9 | P.J.E.Peebles. | 63.6 | J.L.Weiland. | 11.2 |
| 27 J.Polchinski. | 372.8 | A.O.Starinets. | 62.7 | Benjamin.Gold. | 10.9 |
| 28 C.N.Yang. | 370.4 | N.Seiberg. | 60.4 | M.Limon. | 10.8 |
| 29 A.Strominger. | 365.0 | A.D.Martin. | 59.7 | E.Gross. | 10.6 |
| 30 Abdus.Salam. | 338.9 | N.Nekrasov. | 59.4 | M.Herquet. | 10.4 |

Table 10: Authors sorted according to their CitationCoin \( \wp_A \).
becomes a more robust measure: citations from paper \( p' \) are weighted by its rank \( R_{p'} \), giving relatively less weight to ‘below average’ papers that sometimes contain many self-citations. Still, many below average papers can sum up to a significant total rank.

One can optionally count citations from published papers only, ignoring citations from unpublished papers. However this choice discards information and good unpublished papers (some well respected authors don’t publish their papers).

Removing all self-citations is, by itself, an arbitrary choice. Furthermore it can be implemented in different ways, for example removing citations from co-authors. Removing only citations of an author to itself amounts to set to zero the diagonal elements of the citation matrix \( N_{A'\rightarrow A}^{\text{cit}} \), reducing its \( N^2 \) entries to \( N(N-1) \).

This does not protect from ‘citation cartels’. A second step in this direction consists in removing citations exchanges \( A \rightarrow A' \) and \( A' \rightarrow A \) between all pairs of authors \( A,A' \). This amounts to subtract the symmetric part of the \( N_{A'\rightarrow A}^{\text{cit}} \) matrix, reducing its entries to \( N(N-1)/2 \).

A third step is removing citations exchanges \( A \rightarrow A' \), \( A' \rightarrow A'' \) and \( A'' \rightarrow A \) among triplets of authors \( A,A',A'' \). A fourth step is removing all quadruplets etc.

A combinatorial computation shows that, after removing all possible cartels, only \( N-1 \) entries remain. They can be described by \( N \) numbers \( \mathcal{C}_A \) that sum to 0. The meaning of \( \mathcal{C}_A \) can be intuitively understood by viewing \( N_{A'\rightarrow A}^{\text{cit}} \) as the total amount ‘paid’ by \( A' \) to \( A \), and \( N_{A\rightarrow A'}^{\text{cit}} \) as a matrix of transactions. Then the physical quantity unaffected by cartels is the net amount ‘owed’ by each author \( A \):

\[
\mathcal{C}_A = \sum_{A'} (N_{A'\rightarrow A}^{\text{cit}} - N_{A\rightarrow A'}^{\text{cit}}).
\] (15)

Citations are treated like money, because money is the solution to a mathematical problem: subtracting all possible citation cartels is equivalent to count citations received as positive, and citations given as negative. In doing this we proceed as described above: we actually count individual citations (‘icit’): shared between co-authors, and divided by the number of references of each paper. Then the price paid is the total number of papers written, independently of their number of references. The CitationCoin of authors, \( \mathcal{C}_A \), can be written in terms of the CitationCoin of their papers, \( \mathcal{C}_p \):

\[
\mathcal{C}_A = \sum_{p \in A} \mathcal{C}_p N_{\text{aut}}^p, \quad \mathcal{C}_p = N_{\text{icit}}^p - 1.
\] (16)

A paper has \( \mathcal{C}_p < 0 \) when it receives a below-average number of individual citations.$^{13}$ Authors get a large ‘CitationCoin’ by writing papers with above-average quality, especially excellent papers. The CitationCoin rewards both quality and quantity, unlike the ‘average number of citations’, which is maximised by writing very few highest-quality papers.

$^{13}$As time is needed to accumulate citations, this metrics penalises recent papers and young authors. This can be bypassed defining a \( \mathcal{C}^+ \) metric which discards ‘negative’ papers with \( \mathcal{C}_p < 0 \) and sums the contributions of ‘positive’ papers. Only the best papers contribute to \( \mathcal{C}^+ \), analogously to what happens in the \( h \)-index. Unlike the \( h \)-index, \( \mathcal{C}^+ \) also rewards excellent papers.
Table 10 lists authors according to their ‘CitationCoin’ $\mathcal{C}_A$. Authors that scored highly in previous ranks by writing many papers with low impact have now disappeared, and some of them actually got a negative $\mathcal{C}_A$.

Finally, from the database we can extract detailed reports about the metrics of each author: $N_{\text{pap}}, N_{\text{cit}}, h, N_{\text{ipap}}, N_{\text{icit}}, R, \mathcal{R}, \mathcal{C}$, their time evolution, the scientific age, the percentage of given and received self-citations, the topics studied, the main collaborators, who the author cites most, who cites the author most, etc. Similarly, these informations can be extracted and compared for a group of authors, for instance those that applied to some academic position.

4.5 Author metrics: correlations

The right panel of fig. 6 show the correlations among the metrics for authors within the whole INSPIRE database. The metrics are: number of papers $N_{\text{pap}}^A$, number of citations $N_{\text{cit}}^A$; $h$-index squared $h^2_A$; number of individual papers $N_{\text{ipap}}^A$, number of individual citations $N_{\text{icit}}^A$, PaperRank of authors $R_A$, AuthorRank of authors $\mathcal{R}_A$; citation coin $\mathcal{C}_A$. We consider the square of the $h$ index since this is known to be almost fully correlated (0.99) with the number of citations [4]. From the table we see that our metrics for authors differ strongly from the traditional ones, and also mildly different between them. The main difference arises because of the difference between experimentalists and theorists, so that the metrics become more correlated if restricted within each group. However, the combined effect of dividing by the number of references of the citing paper and the number of co-authors of the cited one makes our new bibliometric indicators fairly uncorrelated from the existing ones.

5 Rankings groups

Our metrics respect sum rules and thereby allow to define metrics for groups by simply summing over their members. Furthermore, the main property of the CitationCoin holds not only for authors of a set of papers, but for any group: it cannot be increased through internal citations. In order to show illustrative results we mostly use the individual number of citations as the metric to rank groups. One could equally apply any other metric discussed before.

5.1 Ranking institutions

We share the number of citations received by each paper between the institutions $I$ of its authors with weights $p_I$ that sum to 1. The weights are computed by first sharing equally each paper between its $N_{\text{aut}}$ authors, and next between the affiliations of each author. Thereby each author $A$ contributes to institute $I$ as $1/(N_{\text{aff}}^A N_{\text{paut}}^A)$, where $N_{\text{aff}}^A$ is the number of affiliations of author $A$, that include institute $I$. For example, this paper is attributed 5/12 to CERN, 1/4 to INFN-Genova, 1/6 to INFN-Pisa, 1/6 to Pisa U. When some affiliation is missed by InSpire, we renormalise the $p_I$ such that they sum to 1. Next, we sum over all papers $p$ (optionally restricting to recent papers, if one wants to evaluate the present situation, rather than the
Table 11: Institutions listed according to their contribution to fundamental physics (as defined by INSPIRE) quantified as the number of individual citations received by their affiliates as defined in eq. (17). Left: all time. Right: recent, and split within arXiv categories. The top ten in each category are highlighted.
all-time record). In formulæ, the number of individual citations received by institute $I$ is:

$$N_{cit}^I = \sum_p p_{pI} N_{pI}^{cit}, \quad p_{pI} = \frac{1}{N_{paut}} \sum_{A \in I,p} \frac{1}{N_{aff}^A}. \quad (17)$$

In the left column of table 11 we list the institutes that most contributed to fundamental physics, according to the whole INSPIRE database. In order to focus on the present situation, in the second column we restrict to recent papers, written after 2010. The top positions are occupied by research institutions rather than by teaching institutions. In the right columns of table 11 we show the contributions within each main arXiv category: the best institutions strongly depend on the sub-field of interest. This means that generic rankings (e.g. at the faculty level) are not much useful for authors interested in finding the most active institutions within their specific sub-fields.

Concerning the time evolution, we compute the percentage of individual citations received by papers written within any given year by each institute. The black curve in fig. 9 shows the time-dependence of the contribution of the top institution, CERN. It reached a maximum around 1965 (12% of world-wide individual citations to 1965 papers have been given to CERN authors) and declined stabilising to $\approx 2\%$. All main historical institutions show a similar trend, due to the fact that, around 1970, fundamental physics was concentrated in a few institutions, and later became more distributed (especially in theoretical physics). Half of the impact, quantified using individual citations, was produced in the 12 top institutions in 1970, in 22 in 1980, 42 in 1990, 80 in 2000 and 160 now. As a consequence the relative impact of the top institutions declined. Among the new institutions, Perimeter and IPMU reached very high positions.

The list in table 11 highlights the institutes with the most productive authors in recent times (often young authors). The list in table 12 highlights institutes with the most productive affiliates, as instead quantified by their all-time biblio-metric rankings. Table 12 is produced as follows. Since no list of present affiliates is available, we use the declared affiliations of authors that wrote at least one paper in the last year, 2017. Authors with $N_{aff}$ affiliations are assigned with fraction $1/N_{aff}$ to each affiliation. When this number differs in recent papers, we average over them respecting sum rules: each authors is affiliated to various institutions with percentages that sum to one. The average suppresses minor mistakes/missing data in INSPIRE. For each institute $I$, we obtain a list of active affiliates with their percentages. Summing over these percentages we determine the number of ‘individual authors’ $N_{iaut}$ affiliated to each institution, shown in the 3rd column of table 12. Next, summing over all affiliates using the same weights, we compute the total biblio-metric ranking of all authors in each institute. In column 4 we show the all-time number of individual citations, in column 5 the PaperRank (section 3.1) and in column the AuthorRank (section 4.3). The latter 3 columns actually show the world percentage of each metric in the various institutes: about 2% of researchers that most contributed to fundamental physics can be found at IAS, or at CERN.
Table 12: Institutes listed according to all-time bibliometric ranks of their last-year affiliates. For each institute, we show the number of individual authors active in the last year, and the sum of their bibliometric indicators: number of individual citations, rank based on papers, rank based on authors.

5.2 Ranking towns

Sometimes multiple institutes are located nearby, and what matters is their total. We group together institutes closer than about 30 km. In fig. 7 we show a map with the places that mostly contributed to fundamental physics: each contribution is plotted as a circle with area proportional to the number of individual citations received by their papers written after 2000, and color proportional to the contribution to experiment (green), theory (red), astro-cosmology (blue) respectively. We focus on a relatively recent period, such that the map photographs the present situation.

Similar maps can be computed for any given sub-topic or region. For instance, fig. 8 shows the same map separated according to papers published within the main arXiv categories, and restricted to Europe.

5.3 Ranking countries and continents

We rank a country or continent $C$ by summing the ranks of all institutes located in $C$. We apply this to the number of individual citations:

$$N_{\text{cit}}^C = \sum_{I \in C} N_{\text{cit}}^I .$$


In fig. 9 we show the time evolution of the impact of papers written in main representative countries within each year. The impact is quantified as the percentage with respect to the world total, in order to factor out the reduced number of citations of recent papers. USA is the main country, but declining (from 70% around 1950 to 25% now); European countries are now stable or slightly growing; China is growing. In the right panels of fig. 9 the time evolution of the percentage contribution of each country is shown separately, after the advent of arXiv, within the main fields: experiment, theory, astro-cosmology.

Fig. 10 shows the analogous plot for continents. We see that WW2 lead to the decline and to the later recover of European physics, which returned to be the main actor around 2000. The decline of Asia around 1985 is due to the fall of Soviet Union (Mathematica geographic tools assign all Russia to Asia); the present rise of Asia is mostly due to China.

Fig. 11 shows the ratio between the number of individual citations and the population of countries. What is the reason of the huge variation, by about 7 orders of magnitude? In fig. 12 we show the ratio between the number of individual citations and gross domestic product: the spread between different countries gets only slightly reduced. GDP is not the main factor that accounts for the large spread.

5.4 Ranking journals
When a paper is published on some journal (referred or not), INSpire provides this information. Table 13 lists journals according to the number of individual citations received by all papers they published in fundamental physics, as included in INSpire. We separately show these
Figure 8: As in fig. 7, restricted to Europe, and showing separately the contributions within main arXiv representative categories.

data for all InSpire, and restricting to articles published after 2000. Fig. 13a shows the time evolution of the percentage number of individual citations received by all papers on selected journals. We see that internet brought a revolution around 2000: the decline of NPB and PLB and the emergence of JHEP as preferred journals.

The 3rd column of table 13 shows a direct measure of ‘quality’: the average number of individual citations per paper, which roughly corresponds to what is known as ‘impact factor’. According to this measure, the top-journals are those that publish reviews.

The 4th column shows a measure of both ‘quality’ and ‘quantity’: the CitationCoin $\mathcal{C}$ of the journal (difference between the number of individual citations and the number of published papers). The top journal according to this measure is PRL. Journals that publish reviews score well, but are limited by their restricted scope. The CitationCoin is negative for journals that publish many papers that do not attract much interest, in particular those that publish conference proceedings. Recent papers tend to have $\mathcal{C} < 0$, as they need time to accumulate citations, so this measure is less significant on short time-scales.

5.5 Ranking genders

We consider 70500 authors indexed in the InSpire database (see appendix A for details). The Mathematica machine learning function `Classify` has been trained to automatically assign to each name a tag: male, female or indeterminate (about 40%, that we ignore). `Classify` sometimes fails (the female with highest PaperRank is Nicola Cabibbo, in 51th position): to
HepNames does not include old authors like Noether, the Curie, Mayer.

Among the authors classified as male or female, 16% of the names in the data-base are female, and receive 8.4% of the individual citations and 5.6% of the rank. Fig. 13 shows that the percentage of individual citations received by female authors is growing and is a factor of 2 higher in sub-fields dominated by large experimental collaborations (where bibliometrics cannot identify individual merit) than in more theoretical fields (where social effects are less important).
Figure 11: Individual citations per country after 2010 divided by population. The top countries are Switzerland, Estonia, Italy, Germany, Belgium.

Figure 12: Individual citations per country after 2010 divided by domestic gross product in dollars. The top countries are Estonia, Slovenia, Armenia, Portugal, Italy.

6 Conclusions

We applied improved bibliometric indicators to the InSpire database, which covers fundamental physics mostly after 1970. Fig. 2 shows some main trends: growing rate of papers, growing number of authors and of references per paper. Fig. 4 shows the health status of main sub-fields.

The metrics that we explore are:

- The number of individual citations $N_{icit}$ — defined as the number of citations divided by the numbers of authors of the cited paper and of references of citing papers — compensates for the recent inflationary trends towards more authors and references.

- The PaperRank $R$, which applies the PageRank algorithm proposed by the founders of
Table 13: We show the number of individual citations ($N_{cit} = \sum N_{cit}/N_{ref}$), the average number of individual citations per paper ($N_{cit}/N_{pap}$) and the citation coin ($C / / = N_{cit} - N_{pap}$) for some top journals. The analysis is restricted to fundamental physics as included in the InSpire database. Left: all time. Right: only papers published after 2000.

| Journal, all InSpire | $N_{cit}$ | $N_{cit}/N_{pap}$ | $C / /$ | Journal, after 2000 | $N_{cit}$ | $N_{cit}/N_{pap}$ | $C / /$ |
|----------------------|-----------|-------------------|---------|---------------------|-----------|-------------------|---------|
| 1 Phys.Rev.D         | 99499     | 1.2               | 16034   | Phys.Rev.D          | 39253     | 0.8               | -10148  |
| 2 Phys.Lett.B        | 73739     | 1.3               | 15559   | Astrophys.J.        | 26089     | 0.7               | -11766  |
| 3 Phys.Rev.Lett.     | 50225     | 2.3               | 33354   | JHEP                | 20448     | 0.8               | -4061   |
| 4 Nucl.Phys.B        | 52499     | 2.0               | 26241   | Phys.Rev.Lett.      | 19248     | 1.6               | 7244    |
| 5 Astrophys.J.       | 43590     | 1.0               | -2010   | Phys.Lett.B         | 15497     | 0.9               | -1963   |
| 6 Phys.Rev.C         | 26335     | 0.7               | -12807  | Mon.Not.Roy.Astron.S| 12469     | 0.5               | -13295  |
| 7 JHEP               | 22172     | 0.9               | -3088   | Phys.Rev.C          | 10979     | 0.6               | -6709   |
| 8 Nucl.Phys.A        | 19774     | 0.6               | -12732  | Astron.Astrophys.   | 9734      | 0.5               | -10228  |
| 9 Mon.Not.Roy.Astron.S| 17011     | 0.6               | -11355  | Nucl.Instrum.Meth.A.| 6891      | 0.6               | -4410   |
| 10 Phys.Rev.         | 15966     | 2.2               | 8848    | Nucl.Phys.B         | 6788      | 0.9               | -476    |
| 11 Nucl.Instrum.Meth.A| 14178     | 0.7               | -5020   | Eur.Phys.J.C        | 5871      | 0.8               | -1847   |
| 12 Astron.Astrophys. | 12488     | 0.6               | -10014  | Class.Quant.Grav.   | 4556      | 0.6               | -3592   |
| 13 Phys.Rept.        | 8691      | 6.2               | 7287    | Nucl.Phys.A         | 3737      | 0.4               | -5609   |
| 14 Eur.Phys.J.C      | 7280      | 0.8               | -1314   | JCAP                | 3537      | 0.6               | -2790   |
| 15 Class.Quant.Grav. | 7256      | 0.7               | -3900   | Astron.J.           | 3159      | 0.7               | -1326   |
| 16 Annals Phys.      | 7059      | 1.8               | 3158    | Astrophys.J.Suppl.  | 2895      | 1.8               | 1243    |
| 17 Commun.Math.Phys.| 6911      | 1.6               | 2571    | JINST               | 2247      | 0.6               | -1378   |
| 18 Z.Phys.C          | 6635      | 1.3               | 1449    | J.Phys.G            | 2246      | 0.5               | -2195   |
| 19 Rev.Mod.Phys.     | 6262      | 4.0               | 4690    | Phys.Rept.          | 2025      | 4.5               | 1574    |
| 20 Astron.J.         | 5258      | 0.9               | -312    | Nature              | 1731      | 2.0               | 881     |
| 21 J.Math.Phys.      | 5122      | 0.6               | -3367   | Astropart.Phys.     | 1716      | 1.1               | 89      |
| 22 Astrophys.J.Suppl.| 4269      | 2.2               | 2317    | Nucl.Phys.B Proc.Sup| 1663      | 0.2               | -7285   |
| 23 Prog.Theor.Phys.  | 4093      | 0.5               | -3431   | Int.J.Mod.Phys.A    | 1631      | 0.2               | -5294   |
| 24 Yad.Fiz.          | 3834      | 0.4               | -5816   | Phys.Rev.ST Accel.Be| 1607      | 0.7               | -681    |
| 25 Int.J.Mod.Phys.A  | 3812      | 0.4               | -6349   | PoS                 | 1533      | 0.08              | -17873  |
| 26 Nature            | 3616      | 2.3               | 2014    | AIP Conf.Proc.      | 1407      | 0.07              | -17722  |
| 27 JCAP              | 3537      | 0.6               | -2791   | Comput.Phys.Communn.| 1355      | 1.6               | 525     |
| 28 Nucl.Phys.B Proc.Sup| 3414      | 0.2               | -12593  | Rev.Mod.Phys.       | 1349      | 2.8               | 868     |
| 29 Mod.Phys.Lett.A   | 3240      | 0.4               | -5843   | J.Phys.A            | 1196      | 0.3               | -2850   |
| 30 Comput.Phys.Communn.| 3195      | 1.9               | 1550    | Eur.Phys.J.A        | 1149      | 0.4               | -1775   |

Google to the citation network among papers and that approximates a physical observable: how many times a paper is read.

- The AuthorRank $\mathcal{R}$ which applies the PageRank algorithm to the citation network among authors.

- The CitationCoin $\mathcal{C}$ equal to the difference between the number of received and given individual citations. By treating citations like money, it cannot be increased by self-citations and circular citations.

An important feature of all these metrics is that they can be computed in practice. In section 3 we apply these metrics to papers. Table 1 shows the traditional list of all-time top-cited (highest number of citations) papers. It can be compared with table 2, which shows top-ranked papers (papers with highest PaperRank, namely citations weighted proportionally
to the rank $R$ of citing papers): the PaperRank retrieves old famous papers with relatively few citations. When applied to the time-ordered citation network, the PageRank reduces to a (weighted) counting of citations-of-citations described in section 3.3. Thereby, when restricted to recent papers, the PaperRank is dominated by the number of individual citations. Next, table 3 shows top-referred papers, where citations are weighted proportionally to the all-time AuthorRanks $R_A$ of citing authors. The AuthorRank identifies the recent papers that most attract the attention of notable older authors. Finally the left panel of figure 6 shows the correlations among indicators for papers, showing how our proposed metrics are fairly independent from the traditional ones and among each others.

In section 4 we apply the new metrics to authors. Traditional metrics shown in table 6 are dominated by experimentalists who write more than 100 papers per year in collaborations with 3000 authors. Considering instead the number of individual citations, the list in table 7 becomes dominated by theorists, especially those very active in relatively recent times. Restricting to recent papers, the list includes some authors from fields that tend produce many publications (tens per author per year). Less surprisingly, the list includes authors who produce useful tools for collider experiments, which are presently very active. Ranking authors through their PaperRank, the all-time list in table 8 is dominated by theorists (such as Weinberg, Schwinger, Feynman, Gell Mann) that produced seminal papers after INSPiRE started, despite the overall rate of papers and citations was a factor of few smaller than now (fig. 3). On the other hand, the recent-time list does not change significantly: the PaperRank is strongly correlated to the number of individual citations, becoming a better metric only on longer time-scales.

Due to this limitation, we developed the AuthorRank. Table 9 shows the result: authors such as Dirac and Einstein now appear in the top of the list, despite having few papers with few citations. The right columns of table 9 shows recent authors listed weighting citations according to the all-time AuthorRank of citing authors.
Table 10 lists authors according to their CitationCoin \( \mathcal{C} \): this metric rewards authors who attract the interest of others by writing above-average papers, and penalising those that write many below-average papers (or many recent papers, as papers need decades to be recognized in terms of number of citations). The right columns of table 10 again restrict the list to recent times.

The right panel of figure 6 shows the correlations among indicators for authors. The metrics we propose are fairly uncorrelated with traditional metrics and among each other.

Our metrics respect sum rules (their total is not inflated adding more authors or more references) and are intensive: this means that groups can be ranked summing over their members. In section 5.1 we discussed the institutions that most contribute to fundamental physics, or that contain the authors that most contributed. In section 5.2 we grouped nearby institutes, providing maps of towns most active in fundamental physics. The same is done in section 5.3 for countries and continents: in view of the large statistics we also show the time evolution of their percentage impact. In section 5.4 we compute which journals publish the most interesting results in fundamental physics, again showing the time evolution. Finally, in section 5.5 we compute the percentage contribution of (fe)male authors, and its time-evolution within the main arXiv categories.

The different metrics that we propose give different informations on each author, providing together a more complete view. Paperscape [39] extracts information from arXiv and provides a very useful visualisation of the citation graph among papers, and of the contribution of some authors (those with unique names). It would be interesting to run the open-source Paperscape code on the graph of individual citations among paper and authors extracted from the InSpire database. Our indices could also be implemented in databases that index citations, such as InSpire, in order to offer authors (institutes/journals/group) profiles with a larger variety of information, able to give at one glance a much deeper and wider panoramic of each author (institute/journal/group).

Several of our results with complete tables are available at the webpage [35].

Repetita ad nauseam iuvant  Technical details and limitations are described in appendix A. We repeat the main caveats of our analysis: ‘fundamental physics’ here means ‘as included in the InSpire database’; we do not correct for mistakes in InSpire (anyhow more accurate than commercial databases); etc etc. Omissions should be addressed to feedback@inspirehep.net or trough the on-line forms on InSpire; we will update our results in some future. We just computed and showed results, avoiding comments. We hope that authors, journals, fields, institutes, towns, countries, continents and genders will understand that we cannot repeat all caveats in all results. In particular we said nothing about why women and North Korea are ‘under-represented’.

Acknowledgments

R.T. thanks Roberto Franceschini and Christoffer Petersson for discussions that stimulated this work. We thank Roberto Franceschini for participation in the first stage of this project. We are grateful
to the InSpire team, and especially to Jacopo Notarstefano, for support and help with the InSpire database.

A Details about the dataset

Any large database contains a small fraction of incomplete/inconsistent information, which may affect any algorithmic study of the data in a variety of ways [40]. The InSpire database is extremely curated. References are covered with an accuracy at the % level, typically better than big private databases such as Scopus [41] or WebOfScience [42], and comparable to Google Scholar [13].

We obtained the InSpire database in the form of a ‘dump’ file form ref. [43]. InSpire maps papers, authors, institutes (affiliations), and journals to record IDs (integer numbers) thereby addressing the problem of name disambiguation [36].

The extraction of an accurate date for each paper from the database suffers from some uncertainty. There are several available dates: the date when the paper was added to the database, a preprint date, often, but not always, corresponding to the arXiv preprint (when available), a publication date (if the paper has been published on a journal), and an “earliest” date, representing the first available date (not always present). Moreover, month information is typically available only for arXiv papers and some published ones. In general we estimate our uncertainty on the extracted dates at the percent level, in the sense that dates are accurately extracted (at least the year) for about 99% of the papers. Given this uncertainty our sample of papers consists of 1275708 papers from 1230 [44] to 31 December 2017.

Another source of uncertainty comes from the author list of each paper. Some papers carry an empty list of authors. This can have different reasons. For instance, only the name of the collaboration is available for experimental papers (mainly conference notes) indexed from the CDS database [11]. All these papers are included in our analysis for what concerns papers, but do not contribute to the metrics for authors. Similar problems extends to institutions and journals.

One more source of uncertainty is the extraction of references from papers, needed to produce a citation map. This can be very simple when a bibtex or xml bibliography file is attached to the paper, but can become an extremely complicated task for papers where only a pdf, sometimes produced from a scanned paper, is present. InSpire uses state-of-the-art technology for reference extraction [45,46], which is mainly automatic, with human supervision only in case of errors and inconsistencies. Despite the advanced technology for reference extraction, not all references are correctly extracted.

There are different kind of problems so that a reference can: simply be missed, be recognised incompletely, be misidentified with another one, be assigned to an inexistent paper ID, point out of the database (in which case it is counted in the number of references, but not indexed), or point to a later paper (“acausal” reference). All these effects are observed in the database. While some are simple mistakes, or typos in the ID of the paper, the last could be a real effect (below 1%), with the reference appearing in a subsequent version of the paper with no available information on the dates of the different versions. Since “acausal” citations can generate anomalies in the computation of the rank, we deleted them from our dataset. However, since dates are extracted with an accuracy that is often of one year, we still consider causal all the references to papers with the same date.

In any case, especially because of references pointing outside the database, the number of references indexed for any given record is a better estimate of the actual number of references than the one obtained by summing the indexed ones. Only when computing PaperRanks and AuthorRanks, the number of references is defined equal to the number of indexed references, which is needed to correctly
normalize the transition matrix defining the graph. Given that InSpire is complete only after 1970, this means that references of older papers are over-attributed to those old notable papers that happen to be in InSpire.

In summary, the dataset consists of 1275708 papers, 70388 indexed authors, 7517 institutes, 2508 journals, 30029298 references (of which 21418608 indexed). We compute citations directly from references and do not use citation information from the InSpire database.

Concerning the information on the arXiv database used in figure 1, we imported all records using the arXiv API [37]. All other information on arXiv papers and categories has been obtained from the InSpire database. The full list of arXiv categories and subject-classes can be found in ref. [37].

All indices discussed in this paper can be computed in one hour on a laptop, apart for the Author Rank, which needs a large not very sparse matrix: \( \sim 7 \cdot 10^4 \times 7 \cdot 10^4 \) with about \( 5 \times 10^8 \) non-vanishing entries.

References

[1] M. J. Kurtz and E. A. Henneken, “Measuring metrics - a 40-year longitudinal cross-validation of citations, downloads, and peer review in astrophysics”, *Journal of the American Society for Information Science and Technology* 68 (2017) 695–708, arXiv:1510.09099 [ADS].

[2] E. A. Henneken and M. J. Kurtz, “Usage Bibliometrics as a Tool to Measure Research Activity”, 2017, arXiv:1706.02153 [ADS].

[3] M. J. Kurtz, “Comparing People with Bibliometrics”, 2017, arXiv:1707.09955 [ADS].

[4] J. Hirsch, “An index to quantify an individual’s scientific research output”, *Proc. Nat. Acad. Sci*. 46 (2005), physics/0508025 [InSpire].

[5] “High-Energy Physics Literature InSpire Database”.

[6] INSPIRE Collaboration, A. Holtkamp, S. Mele, T. Simko, and T. Smith, “Inspire: Realizing the dream of a global digital library in high-energy physics”, 2010 [InSpire].

[7] R. Ivanov and L. Raac, “Inspire: A new scientific information system for hep”, *J.Phys.Conf.Ser.* 219 (2010) [InSpire].

[8] J. Klem and J. Iwaszkiewicz, “Physicists get inspired: Inspire project and grid applications”, *J.Phys.Conf.Ser.* 331 (2011) [InSpire].

[9] J. Martin Montull, “Inspire: Managing metadata in a global digital library for high-energy physics”, *Commun.Comput.Inf.Sci.* 240 (2011) 269–274 [InSpire].

[10] “ADS - The SAO/NASA Astrophysics Data System”.

[11] “CDS - CERN Document Server”.

[12] “arXiv.org”.

[13] “Google Scholar”.

[14] “Microsoft Academic”.

[15] “DBLP - Computer Science Bibliography”.

[16] “ACM Digital Library”.

[17] “PubMed”.

[18] “MathSciNet - Mathematical Reviews”.

[19] “CiteSeerX”.

[20] “Semantic Scholar”.

[21] “RePEc - Research Papers in Economics”.

[22] “IEEE Xplore Digital Library”.

[23] “zbMATH the first resource for mathematics”.

[24] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine”, *Comput. Netw.* 30 no. 1-7, (1998) 107–117 [Semantic Scholar].

[25] S. Brin, R. M. Larry Page, and T. Winograd, “The PageRank citation ranking: Bringing order to the web”, *Stanford InfoLab*, (1999) [Semantic Scholar].

[26] R. Tanase and R. Radu, “The Mathematica of Web Search”.

[27] A. Rajaraman and J. D. Ullman, “Mining of massive datasets”, *Cambridge University Press* (2009) [Semantic Scholar].
[28] P. Chen, H. Xie, S. Maslov, and S. Redner, “Finding scientific gems with google”, J.Informet. 1 (2007) 8–15, physics/0604130 [InSpire].

[29] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles, “Co-ranking authors and documents in a heterogeneous network”, Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007), October 28-31, 2007, Omaha, Nebraska, USA (2007) 739–744 [Semantic Scholar].

[30] N. Ma, J. Guan, and Y. Zhao, “Bringing pagerank to the citation analysis”, Inf. Process. Manage. 44 no. 2, (2008) 800–810 [Semantic Scholar].

[31] Y. Ding, E. Yan, A. R. Frazho, and J. Caverlee, “PageRank for ranking authors in co-citation networks”, JASIST 60 no. 11, (2009) 2229–2243, arXiv:1012.4872 [Semantic Scholar].

[32] D. A. Bini, G. M. Del Corso, and F. Romani, “Evaluating scientific products by means of citation-based models: a first analysis and validation.”, ETNA, Electron. Trans. Numer. Anal. 33 (2008) 1–16 [zbMATH].

[33] G. M. del Corso and F. Romani, “Versatile weighting strategies for a citation-based research evaluation model.”, Bull. Belg. Math. Soc. - Simon Stevin 16 no. 4, (2009) 723–743 [zbMATH].

[34] D. Bini, G. M. D. Corso, and F. Romani, “A combined approach for evaluating papers, authors and scientific journals”, J. Comput. Appl. Math. 234 (2010) 3104–3121 [Semantic Scholar].

[35] “Biblioranking Fundamental Physics Web-page”.

[36] M. Golosovsky and S. Solomon, “Growing complex network of citations of scientific papers - measurements and modeling”, Phys. Rev. E 95 no. 012324, (2017), arXiv:1607.08370 [Semantic Scholar].

[37] “arXiv.org API”.

[38] S. Lehmann, B. Lautrup, and A. Jackson, “Citation distributions in high-energy physics”, Phys. Rev. E 68 (2003), physics/0211010 [InSpire].

[39] “PaperScape”.

[40] G. M. D. Corso and F. Romani, “A multi-class approach for ranking graph nodes: models and experiments with incomplete data”, Inf. Sci. 329 (2016) 619–637 [Semantic Scholar].

[41] “Scopus”.

[42] “WebOfScience”.

[43] “Inspire Dump”.

[44] J. de Sacrobosco, “Tractatus de sphaera”, 1230 [InSpire].

[45] “Refextract reference extractor”.

[46] “Grobid reference extractor”.

35