Implementation of the PaperRank and AuthorRank indices in the Scopus database

Original Citation:

Availability:
This version is available at: 2158/1241375 since: 2021-09-08T09:15:46Z

Published version:
DOI: 10.1016/j.joi.2021.101206

Terms of use:
Open Access
La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:

(Article begins on next page)
Implementation of the *PaperRank* and *AuthorRank* indices in the Scopus database

Pierluigi Amodio*  Luigi Brugnano†  Filippo Scarselli‡  
August 3, 2021

**Abstract**

We implement the *PaperRank* and *AuthorRank* indices introduced in [Amodio & Brugnano, 2014] in the Scopus database, in order to highlight quantitative and qualitative information that the bare number of citations and/or the h-index of an author are unable to provide. In addition to this, the new indices can be cheaply updated in Scopus, since this has a cost comparable to that of updating the number of citations. Some examples are reported to provide insight in their potentialities, as well as possible extensions.

**Keywords:** bibliometric index, *PaperRank*, *AuthorRank*, Scopus database, number of citations, h-index.

1 Introduction

This paper deals with the interesting question of ranking authors and scientific publications. This is a very important and debated topic, especially since the coverage of bibliometric databases has made it possible to correctly analyze the scientific path of each author, at least for those of the last 50
years [9, 13, 26]. Today, citation analysis represents a reliable formal tool for quantitative scientific assessment.

For our research we use the Scopus database [30] which assigns to each author an identifier (the Scopus A.Id) in order to count, on a Scopus dedicated page, the list of the authored papers and the references therein. This allows us to immediately know, not only the number of papers, but also the number of citations of each author in the database. To these values Scopus adds, thanks to its simplicity of calculation, the h-index [14], that is, a value equal to $k$ if the author authored $k$ papers with at least $k$ citations.

These parameters, although widely used, have drawbacks which lead to results that are sometimes not meaningful (see, e.g., [7, 17, 21] published this year (2021) and Section 4). For this reason, many other indices have been proposed (see, for example, [4]). Some of them (but this is only a very short list) just follows from the h-index: m-index [14], g-index [10, 11], e-index [27], $h^2$-index [23]. Others, as the i10-index (used by Google Scholar [29]) which counts the number of papers with more than 10 citations, derives from a different observation of the data. Anyway, none of these indices takes into account the fact that, for example:

- papers from different subject areas usually use bibliographies with quite different lengths;
- within the same subject area, the number of citations increases over time;
- review papers have very long reference lists while very targeted research works usually have a small bibliography;
- self citations may be a significant part of the citations.

For this reason other indices have been introduced which, however, require a more complex processing of the data contained in the database. In this regard, an algorithm, that can certainly be taken into consideration as a reference, is the PageRank of Google [22] which, when a Google search is performed, assigns a score to each page and then provides the user with a ranking of the most interesting pages [28]. The PageRank algorithm is based on the number of incoming links and the weighting of the linking page. Irrespective of the content of a given page, this latter is better evaluated when important pages link to it. The ranking of a site is therefore recursive to the
assessment of the linking page (the PageRank depends on the number of pages linking to it, and the quality of these pages).

Bringing this idea back to our problem, the idea is that the importance of a paper depends not only on the citations received, but also on the importance of the papers that make these citations. Several authors have exploited this idea to define new indices, for example P-Rank [20], AuthorRank [19], Article Rank [18], Y-factor [3], CiteRank [24] and Eigenfactor [2, 25]. The problem is that the calculation of these indices requires a high computational cost, and an update of the reference database (which happens at least weekly on Scopus) requires the complete recalculation of the computed indices. For this reason, the numerical tests carried out with these indices, always refer to much smaller databases, or analyze journals of a given subject area [8, 12, 20].

In [1], a mathematically correct simplification of the PageRank model has been proposed, which requires a computational cost comparable to that of the h-index. For this reason, it could be easily implemented in the Scopus database. A further feature of this new index is that it is additive, and still preserving the main properties of the Google Rank. This means that it can be easily updated, by taking into account possible new citations.

With these premises, the structure of the paper is as follows: in Section 2 we recall the main facts about the AuthorRank and PaperRank presented in [1]; in Section 3 we introduce a web page that has been set up for computing these new parameters by means of a few Scopus API keys [31], and in Section 4 we provide some simple examples that highlight their potential; finally, in Section 5 some new indices are defined, that could be derived (for free) to provide further insight.

2 Materials and methods: definition of PaperRank and AuthorRank

The idea behind the definition of PaperRank in [1] derives, as mentioned above, from that of the Google PageRank [22] which, in our case can be rewritten and summarized in this way:

“an important paper is cited by important papers”.

From a mathematical point of view, starting from a database containing the papers, each with its own references, it is possible to obtain the impor-
tance of a paper by calculating the eigenvector $v$ associated with the unit eigenvalue of the matrix $S = LF^{-1}$, where $L$ is the citation matrix (of size equal to the size of the database)

$$L = (\ell_{ij}), \quad \ell_{ij} = \begin{cases} 1 & \text{if paper } j \text{ cites paper } i, \\
0 & \text{otherwise}, \end{cases} \quad (1)$$

and $F$ is a diagonal matrix, whose diagonal entries are defined as

$$f_j = \sum_i \ell_{ij}. \quad (2)$$

We observe that $f_j$ is nothing but the number of references in the bibliography of the paper $j$. For the sake of simplicity, we assume that $f_j > 0$, even though this will be not strictly necessary, as explained below (see [1] for more details). By setting, hereafter, $e$ the vector of length $n$ with all entries equal to one, one easily realizes that

$$e^\top S = e^\top LF^{-1} = (f_1, \ldots, f_n)^\top F^{-1} = e^\top,$$

so that $\lambda = 1$ is an eigenvalue of matrix $S$. Similarly, one has that $\|S\|_1 = 1$, where we recall that the 1-norm of a matrix is the maximum among the sums of the absolute values of the entries on each row (in particular, the sum on each row of $S$, which is a non-negative matrix, is 1). Consequently, since the spectral radius of $S$ is less than or equal any natural norm, the unit eigenvalue is the one of maximum modulus.

The calculation of the eigenvector $v$ requires an information of global nature, as in the case of Google PageRank (see, for example, [15, 16] where efficient strategies are implemented to reduce the corresponding computational cost). In addition, any update of the database would require a complete recalculation of the vector $v$. For this reason, in [1] the following, much cheaper, approximation was proposed,

$$v_0 := Se = LF^{-1}e. \quad (3)$$

We observe that the vector $v_0$ can be also regarded as the first step of the power method, applied for computing the right eigenvector of the unit eigenvalue of $S$, starting from $e$.

This is the best approximation we can derive by using only a local information: in fact, its computational cost equals that for computing the bare
number of citations (i.e., the vector $Le$) and turns out to be additive, that is, an update of the database requires to update the approximation only taking into account the new entries (i.e., the newly added citations).

Hereafter, we shall refer to the vector $v_0$ as the PaperRank $[1]$. Moreover:

- we shall denote by $n$ the number of papers in the database (i.e., matrix $L$ is of size $n \times n$);
- for each paper $P_j$ in the database, we set:
  \begin{itemize}
  \item $\#Auth(P_j)$ the number of authors of the paper,
  \item $\#Ref(P_j)$ the number of references in its bibliography (i.e., $f_j$),
  \item $n_j$ the number of papers citing it.
  \end{itemize}

Let $c_{1,j}, \ldots, c_{n_j,j}$ be the indices of the papers citing paper $P_j$. Then, according to (1)–(3), we define:

$$PaperRank(P_j) = \frac{1}{\#Ref(P_{c_{i,j}})} \sum_{i=1}^{n_j} \frac{1}{\#Ref(P_{c_{i,j}})}, \quad j = 1, \ldots, n.$$  \hfill (4)

From this definition we also get that of AuthorRank. In more detail:

- let $m$ be the number of all authors of all papers in the database;
- for each author $A_k$, we set $m_k$ the number of her/his papers.

Let $b_{1,k}, \ldots, b_{m,k}$ be the indices of the papers written by the author $A_k$. Then, we set:

$$AuthorRank(A_k) = \frac{PaperRank(P_{b_{j,k}})}{\#Auth(P_{b_{j,k}})} \sum_{j=1}^{m_k} PaperRank(P_{b_{j,k}}), \quad k = 1, \ldots, m.$$  \hfill (5)

**Remark 1** In other words, the PaperRank of each paper is equally divided among all its authors. Therefore, it is easy to show that (see Theorem 2 in [1]):

$$\sum_{j=1}^{n} PaperRank(P_j) = \sum_{k=1}^{m} AuthorRank(A_k).$$

It is worth mentioning, however, that any discrete probability distribution could in principle be used (taking into account, as an example, the order of the authors), even though we shall not consider this possibility, hereafter.
Example 1 To better explain how the PaperRank and the AuthorRank can be easily updated, let us assume that a given author, having AuthorRank = 10, receives a new citation concerning one of her/his paper, having PaperRank = 2, written together with another co-author, and that this citation is done by a paper with 5 references. As a result:

- the AuthorRank is updated as \[ \text{new-AuthorRank} = 10 + \frac{1}{2 \times 5} = 10.1; \]
- the PaperRank is updated as \[ \text{new-PaperRank} = 2 + \frac{1}{5} = 2.2; \]

thus confirming the incremental nature of such indices.

3 A web page to compute the AuthorRank

We have developed a web page which is able to query the Scopus database in order to compute the rank of each paper (PaperRank) and, consequently, of each author (AuthorRank). This is possible since in Scopus:

- any author is identified by a unique numeric code (i.e., the Scopus Author Identifier mentioned above, A.Id in short);
- any paper is identified by an analogous alphanumeric string (Electronic Identifier, eid in short), beginning with the characters 2-s2.0-.

Scopus, beyond immediately providing some information about each author (including number of publications and h-index) and publication (for example the bibliography and the number of citations), allows to formulate additional queries through API keys that are well explained starting from the web page [31].

Through these API keys it is possible to retrieve the information needed to calculate the PaperRank and the AuthorRank. In particular, to compute the PaperRank of a given paper:

- the number of the received citations;
- for each citation, the length of the bibliography of the citing paper;
- the PaperRank is then computed according to (4).
Obviously the first value is the only one that can be updated (increased) over time. According to what has been exposed in the previous section, to compute the AuthorRank of a specific author, one has to compute the number of the author’s publications and, for each publication, the corresponding PaperRank, as specified above. Then the AuthorRank is computed according to (5).

Therefore, the computation of the PaperRank requires a number of queries equal to the number of citations of the considered paper. Correspondingly, the computation of the AuthorRank requires a number of queries equal to the number of papers of the authors plus the sum of the queries to compute the PaperRank of each one of such papers. With the obtained informations, both the PaperRank and the AuthorRank are calculated by means of simple weighted averages, according to (4) and (5). Furthermore, since these operations are associative, we stress that updating the AuthorRank and/or the PaperRank only require to consider:

- possible new citations of a given paper;
- possible new papers of a given author.

Actually, the web page allows to find an author starting from his name and/or the A.Id and compute the AuthorRank and the PaperRank of each document to date. Moreover, since all the information allowing the calculation is stored (with the data of computation), it is also possible to dynamically update these indices.

**Remark 2** We would like to emphasize that, should the value of the PaperRank be added to the Scopus record of each paper, then each new citation of that paper would result in an increase in its PaperRank equal to 1 over the length of the bibliography of the paper making the new citation.

Equivalently, any newly added paper in the Scopus database would increase the PaperRank of each paper referenced in its bibliography by 1 over the length of the bibliography itself (in fact, the sum of these updates has to be one).

Similarly, any update in the PaperRank of a given paper, would result in an update of the AuthorRank of its authors equal to the increase of the PaperRank divided by the number of the authors.\(^1\)

\(^1\)We want to emphasize that the web page has been set up for exhibition purposes, and relies on the use of the API keys, which Elsevier has granted us for this research.
4 Results

The aim of this section is to highlight the potential of the PaperRank and the AuthorRank. For this purpose, we shall denote:

1. \#Cit the number of citations obtained by a given paper;
2. \#Pub the number of publications of a given author;
3. \(\sum\text{Cit}\) the sum of citations received by (all papers of) a given author;
4. \(\sum\text{P.R.}\) the sum of the PaperRank of (all papers of) a given author.

The first three parameters are well highlighted in the Scopus database, for any paper/author. However, they may not be sufficient to properly rank them. In particular, the following aspects can be stressed:

a) Authors having a low h-index may have a high AuthorRank.

b) A paper with a higher number of citations than another one, may have a lower PaperRank.

Concerning point a), we report the following two examples:

- the authors of the seminal paper [22], Brin and Page, have a low h-index. In fact, Brin’s h-index is 10. Moreover, Page has two different A.Id in Scopus, one with h-index 3, and one with h-index 1. We computed the corresponding AuthorRank and, in particular, that of Page is nothing but the bare sum of the AuthorRank of his two profiles. As a result, we obtained AuthorRank 221.76 and 166.36 for the two authors, respectively, largely due to the citations of their most important paper [22] and its reprint [6] (having PaperRank 319.37 and 17.05, respectively);

- the two 2013 Nobel Prizes in Physics, F. Englert and P.W. Higgs have a very different h-index, 27 the former and 9 the latter. Nevertheless, \(\sum\text{P.R.}\) of their respective papers (at least, those contained in Scopus) turn out to be comparable (161.31 vs. 182.345), and the AuthorRank are reversed (73.28 vs. 182.24), due to the fact that almost all the papers by Higgs are with only one author.
Concerning point \textbf{b}), we observe that this naturally happens for papers from different subject areas. Nevertheless, this may happen also for papers within the same subject area and by the same author. For example, analyzing the papers with the highest number of citations in the huge bibliography of Karplus (\textit{Author Rank} = 1010.29, see Table 3) it turns out that a work published in 2009 with 34 co-authors has over 4500 citations and a \textit{Paper Rank} = 78.11, while a work (with a single name) published in 1958 with 2600 citations has a \textit{Paper Rank} = 91.14. This is due to the fact that, in recent years, as the h-index has become more and more important for academic evaluations, the bibliography length has grown a lot. Consequently, a citation today generally weighs much less than one obtained several years ago.

To better highlight the features of the new indices, it would be appropriate to compare authors having the same \(\sum\text{Cit}\) but, unfortunately, Scopus, through the Author Search, allows to sort (apart from the useless, for our purpose, alphabetical order) only by the h-index and by the Document Count. For this reason, we now perform some test concerning authors with the same h-index. In particular, Table 1 contains the 31 authors whose name begins with the letters “AL” and having, on 15 January 2021, an h-index = 48. These authors (see Table 1) come from 12 different “Subject Areas” (Scopus identifies 26 of them) which, for the sake of completeness, are fully explained in Table 2.

Analyzing Table 1 we observe that the total number of publication may vary greatly (#Pub ranges from 68 to 352), as well as the total citations (\(\sum\text{Cit}\) ranges from 6133 to 62815). Moreover, while #Pub is not significantly related to our parameters, \(\sum\text{Cit}\) seems roughly proportional to the \(\sum\text{P.R.}\) of the considered author, as is shown in Fig. 1. In this figure, \(\sum\text{Cit}\) is on the abscissa, and \(\sum\text{P.R.}\) on the ordinate (for ease of reading the graph, the point with the largest value has been cut, since it is significantly farther from the others). A dashed line indicates the linear regression line between \(\sum\text{Cit}\) and \(\sum\text{P.R.}\). It can be immediately seen that many points are close to the regression line but some of them are very far. In particular, one author (subject area “Physics and Astronomy”) has a \(\sum\text{P.R.}\) value that is more than double the expected value, while three authors (subject areas “Environmental Science” and “Earth and Planetary Sciences”) have a \(\sum\text{P.R.}\) that is just over half of the expected value: the points corresponding to such authors are marked in Fig. 1 with triangles.

Further, to better understand these results, all the 75 papers having more
than 500 citations by the authors listed in Table 1 have been analyzed in more
details. For each paper, we have computed the corresponding ratio
\[
\rho = \frac{\#\text{Cit}}{\text{PaperRank}},
\]
i.e., the average number of citations needed to increase the PaperRank of the
given paper by one. It turns out that \( \rho \) approximately varies in the interval
[14, 67]:

- the lowest values of \( \rho \) (say, less than 25) are for papers in “Physics and
  Astronomy” and “Medicine” subject areas;

- the highest values (greater than 60) are obtained for papers in the
  “Environmental Science” and “Agricultural and Biological Sciences”
  subject areas.

As is clear, these results are in some agreement with the triangles marked in
Fig.

The results in Fig. 2 are certainly more significant: here the \( \sum \text{Cit} \) value is
on the abscissa while the AuthorRank is on the ordinate (again, the plot does
not include the author with the highest number of citations and AuthorRank
(AuthorRank = 946.37), which would appear far, on the upper-right of the
figure). The other authors have an AuthorRank which varies in a fairly wide
range from 0.49 to 149.06. Half of them have AuthorRank between 35 and
50 (relatively close to the value AuthorRank = h-index = 48 indicated by
the dashed line in the figure), but obviously the authors with values outside
this range are meaningful. In particular, the two authors in the “Physics
and Astronomy” area listed in Table 1 have AuthorRank equal to 0.49 and
119.47, respectively, despite having a comparable \( \sum \text{Cit} \) number (see the
triangles in Fig. 2). This huge difference depends on the fact that, while the
former (an experimental physicist) has publications in collaboration with
over 100 authors, the latter (a theoretical physicist) has publications with a
few co-authors and the papers citing him have very targeted bibliographies
(17 bibliographical items, on average).

Summing up, even though we have randomly chosen the set of authors,
the obtained results allowed to emphasize some remarkable situations:

- the author in Table 1 with the highest AuthorRank (Reka Albert, Author-
  Rank = 946.37, fourth line in the table) presents in Scopus a publica-
  tion with over 20 thousand citations (PaperRank = 822,92) and at
<table>
<thead>
<tr>
<th>Scopus A.Id</th>
<th>Subj. Area</th>
<th>#Pub</th>
<th>(\sum) Cit</th>
<th>(\sum) P.R.</th>
<th>A.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>24782201700</td>
<td>Agr&amp;BioSci</td>
<td>151</td>
<td>11078</td>
<td>219.32</td>
<td>28.59</td>
</tr>
<tr>
<td>7103056456</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>146</td>
<td>8003</td>
<td>163.60</td>
<td>44.98</td>
</tr>
<tr>
<td>7006175921</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>82</td>
<td>9768</td>
<td>199.77</td>
<td>43.42</td>
</tr>
<tr>
<td>7202686127</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>142</td>
<td>62815</td>
<td>2176.80</td>
<td>946.37</td>
</tr>
<tr>
<td>7005804651</td>
<td>Chem</td>
<td>218</td>
<td>7623</td>
<td>134.81</td>
<td>44.94</td>
</tr>
<tr>
<td>7102471934</td>
<td>Chem</td>
<td>170</td>
<td>9038</td>
<td>317.27</td>
<td>108.08</td>
</tr>
<tr>
<td>7005780945</td>
<td>ComputSci</td>
<td>357</td>
<td>9066</td>
<td>246.75</td>
<td>71.42</td>
</tr>
<tr>
<td>57201695156</td>
<td>Earth&amp;Plat</td>
<td>129</td>
<td>9766</td>
<td>165.42</td>
<td>49.54</td>
</tr>
<tr>
<td>7004116785</td>
<td>EnvirSci</td>
<td>226</td>
<td>6133</td>
<td>164.09</td>
<td>59.94</td>
</tr>
<tr>
<td>7004930908</td>
<td>EnvirSci</td>
<td>122</td>
<td>7231</td>
<td>167.43</td>
<td>48.02</td>
</tr>
<tr>
<td>7402266526</td>
<td>EnvirSci</td>
<td>88</td>
<td>18058</td>
<td>292.66</td>
<td>45.62</td>
</tr>
<tr>
<td>7202917464</td>
<td>EnvirSci</td>
<td>97</td>
<td>19084</td>
<td>371.04</td>
<td>149.06</td>
</tr>
<tr>
<td>7006980158</td>
<td>Imm&amp;MicBio</td>
<td>99</td>
<td>9455</td>
<td>182.86</td>
<td>32.82</td>
</tr>
<tr>
<td>6603637799</td>
<td>MaterSci</td>
<td>139</td>
<td>10731</td>
<td>256.17</td>
<td>41.40</td>
</tr>
<tr>
<td>7006852195</td>
<td>Medic</td>
<td>193</td>
<td>7039</td>
<td>257.76</td>
<td>62.05</td>
</tr>
<tr>
<td>7102809705</td>
<td>Medic</td>
<td>210</td>
<td>7247</td>
<td>255.39</td>
<td>37.47</td>
</tr>
<tr>
<td>7005015688</td>
<td>Medic</td>
<td>248</td>
<td>8182</td>
<td>187.65</td>
<td>35.70</td>
</tr>
<tr>
<td>7202794939</td>
<td>Medic</td>
<td>315</td>
<td>8267</td>
<td>258.50</td>
<td>58.89</td>
</tr>
<tr>
<td>7004446309</td>
<td>Medic</td>
<td>316</td>
<td>8780</td>
<td>263.53</td>
<td>56.50</td>
</tr>
<tr>
<td>6701722543</td>
<td>Medic</td>
<td>159</td>
<td>8997</td>
<td>274.68</td>
<td>29.04</td>
</tr>
<tr>
<td>7004046246</td>
<td>Medic</td>
<td>319</td>
<td>9364</td>
<td>306.80</td>
<td>24.56</td>
</tr>
<tr>
<td>7403276785</td>
<td>Medic</td>
<td>278</td>
<td>9522</td>
<td>285.58</td>
<td>52.61</td>
</tr>
<tr>
<td>7006198419</td>
<td>Medic</td>
<td>378</td>
<td>10083</td>
<td>327.92</td>
<td>35.61</td>
</tr>
<tr>
<td>7101861992</td>
<td>Medic</td>
<td>163</td>
<td>11468</td>
<td>347.37</td>
<td>35.47</td>
</tr>
<tr>
<td>7006253731</td>
<td>Medic</td>
<td>301</td>
<td>20342</td>
<td>797.84</td>
<td>39.56</td>
</tr>
<tr>
<td>5572637370</td>
<td>NeuSci</td>
<td>68</td>
<td>6520</td>
<td>117.78</td>
<td>40.89</td>
</tr>
<tr>
<td>7003644890</td>
<td>NeuSci</td>
<td>170</td>
<td>6613</td>
<td>109.35</td>
<td>17.79</td>
</tr>
<tr>
<td>7103174875</td>
<td>Pharm</td>
<td>184</td>
<td>8670</td>
<td>217.59</td>
<td>45.95</td>
</tr>
<tr>
<td>25641028400</td>
<td>Pharm</td>
<td>299</td>
<td>9714</td>
<td>233.45</td>
<td>49.72</td>
</tr>
<tr>
<td>35228729900</td>
<td>Phys&amp;Astr</td>
<td>164</td>
<td>7612</td>
<td>218.08</td>
<td>0.49</td>
</tr>
<tr>
<td>7201985737</td>
<td>Phys&amp;Astr</td>
<td>352</td>
<td>8658</td>
<td>505.31</td>
<td>119.47</td>
</tr>
</tbody>
</table>

Table 1: Sum of PaperRank and AuthorRank for authors with h-index=48 and surname starting with the letters “AL”, along with their number of publications and total citations, grouped by subject area (listed in alphabetical order).
least 6 others with a number of citations greater than 1000. This value
of the AuthorRank, comparable as we will see later, with that of Stephen
Hawking (AuthorRank = 984.60, see Table 3), is not highlighted by the
h-index;

- the 11 authors classified in the Subject class “Medicine” have an Author-
  Rank ranging from about 24.5 to about 62. Also for these authors the
  value $\sum$Cit definitely does not emphasize the obtained ranking. In par-
  ticular, the author with the highest AuthorRank has the smallest $\sum$Cit
  while the author with the highest $\sum$Cit has an AuthorRank smaller
  than 40 (see the squares in Fig. 2).

Fig. 2 also shows that more than 50% of the authors have an Author-
Rank smaller than the h-index (in particular, two of them, corresponding
to the rightmost dots below the dashed-line, having a very high number of
citations). We now further investigate this aspect, by considering some of
the 50 most influential scientists in the world today [32]: they are listed in
Table 3, along with the previously mentioned authors Hawking and Karplus.
In particular, we consider the authors in [32] with $\sum$Cit ranging from 5000
to 25000, according to the Scopus database, resulting into the 16 scientists

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agr&amp;BioSci</td>
<td>Agricultural and Biolog. Sciences</td>
</tr>
<tr>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>Bioch., Genetics and Mol. Biology</td>
</tr>
<tr>
<td>Chem</td>
<td>Chemistry</td>
</tr>
<tr>
<td>ComputSci</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Earth&amp;Plat</td>
<td>Earth and Planetary Sciences</td>
</tr>
<tr>
<td>Engin</td>
<td>Engineering</td>
</tr>
<tr>
<td>EnvirSci</td>
<td>Environmental Science</td>
</tr>
<tr>
<td>Imm&amp;MicBio</td>
<td>Immunology and Microbiology</td>
</tr>
<tr>
<td>MaterSci</td>
<td>Materials Science</td>
</tr>
<tr>
<td>Math</td>
<td>Mathematics</td>
</tr>
<tr>
<td>Medic</td>
<td>Medicine</td>
</tr>
<tr>
<td>NeuSci</td>
<td>Neuroscience</td>
</tr>
<tr>
<td>Pharm</td>
<td>Pharmacol., Toxicol. and Pharmaceut.</td>
</tr>
<tr>
<td>Phys&amp;Astr</td>
<td>Physics and Astronomy</td>
</tr>
</tbody>
</table>

Table 2: Acronyms used by Scopus to denote the considered subject areas.
<table>
<thead>
<tr>
<th>Author Name</th>
<th>Subj. Area</th>
<th>#Pub</th>
<th>$\sum \text{Cit}$</th>
<th>h-ind.</th>
<th>$\sum \text{P.R.}$</th>
<th>Aut.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks II, R.</td>
<td>Engin</td>
<td>302</td>
<td>5674</td>
<td>36</td>
<td>284.94</td>
<td>81.26</td>
</tr>
<tr>
<td>Bonner, J.T.</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>97</td>
<td>5789</td>
<td>32</td>
<td>185.09</td>
<td>96.41</td>
</tr>
<tr>
<td>Goodall, J.</td>
<td>Agr&amp;BioSci</td>
<td>59</td>
<td>6118</td>
<td>35</td>
<td>115.04</td>
<td>27.28</td>
</tr>
<tr>
<td>Higgs, P.</td>
<td>Phys&amp;Astr</td>
<td>19</td>
<td>6466</td>
<td>9</td>
<td>182.34</td>
<td>182.24</td>
</tr>
<tr>
<td>Vestergaard Hau, L.</td>
<td>Phys&amp;Astr</td>
<td>49</td>
<td>6847</td>
<td>18</td>
<td>285.80</td>
<td>74.88</td>
</tr>
<tr>
<td>Bray, D.</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>111</td>
<td>10314</td>
<td>48</td>
<td>268.41</td>
<td>118.59</td>
</tr>
<tr>
<td>Geller, M.</td>
<td>Phys&amp;Astr</td>
<td>216</td>
<td>11993</td>
<td>61</td>
<td>217.32</td>
<td>59.99</td>
</tr>
<tr>
<td>Knuth, D.E.</td>
<td>Math</td>
<td>133</td>
<td>12914</td>
<td>42</td>
<td>701.57</td>
<td>488.41</td>
</tr>
<tr>
<td>Ogawa, S.</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>56</td>
<td>14559</td>
<td>31</td>
<td>413.87</td>
<td>91.84</td>
</tr>
<tr>
<td>Berners-Lee, T.</td>
<td>ComputSci</td>
<td>57</td>
<td>15424</td>
<td>21</td>
<td>753.50</td>
<td>246.84</td>
</tr>
<tr>
<td>Aspect, A.</td>
<td>Phys&amp;Astr</td>
<td>219</td>
<td>15519</td>
<td>51</td>
<td>440.19</td>
<td>214.11</td>
</tr>
<tr>
<td>Penrose, R.</td>
<td>Phys&amp;Astr</td>
<td>133</td>
<td>15945</td>
<td>45</td>
<td>679.65</td>
<td>493.46</td>
</tr>
<tr>
<td>Watson, J.D.</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>128</td>
<td>15998</td>
<td>39</td>
<td>440.19</td>
<td>214.11</td>
</tr>
<tr>
<td>Wilson, E.</td>
<td>Agr&amp;BioSci</td>
<td>228</td>
<td>20495</td>
<td>57</td>
<td>481.95</td>
<td>288.45</td>
</tr>
<tr>
<td>Montagnier, L.</td>
<td>Medic</td>
<td>380</td>
<td>24647</td>
<td>71</td>
<td>760.46</td>
<td>136.71</td>
</tr>
<tr>
<td>Hawking, S.</td>
<td>Phys&amp;Astr</td>
<td>163</td>
<td>41757</td>
<td>75</td>
<td>1339.02</td>
<td>984.60</td>
</tr>
<tr>
<td>Karplus, M.</td>
<td>Bioc&amp;Ge&amp;Mo</td>
<td>812</td>
<td>119539</td>
<td>157</td>
<td>2754.81</td>
<td>1010.29</td>
</tr>
</tbody>
</table>

Table 3: Sum of PaperRank ($\sum \text{P.R.}$) and AuthorRank (Aut.R.) for the authors in [32] with $\sum \text{Cit}$ in the range [5000, 25000], sorted by increasing number of citations.
Figure 1: Sum of PaperRank ($\sum P.R.$) with respect to the total number of citations ($\sum \text{Cit}$), for the authors with h-index=48 listed in Table 1.

listed in Table 3. In Fig. 3, we plot the AuthorRank of each author w.r.t the corresponding h-index. From the figure, one may see that for almost all these authors, the AuthorRank is much greater than the corresponding h-index (see the dashed line representing AuthorRank = h-index): almost 20 times for Higgs who has h-index = 9, more than 10 times for Knuth and Penrose who have AuthorRank $\approx 500$, which are marked with the triangles in Fig. 3.

Incidentally, the most cited papers of Knuth and Wilson have approximately the same PaperRank (179.63 and 175.12, respectively), though having a quite different number of citations (approximately 3800 and 8600, respectively). This means that also papers from authors in different subject areas are more effectively compared.
5 New indices and future work

Although the results obtained in the previous section seem very encouraging, possible requests of different indicators from some subject areas may require a non-standard use of PaperRank and AuthorRank. In fact, the simplicity of their definition could give a simple answer to these questions, too. For example, for some subject areas, where it is usual to have papers with many authors, the partitioning of the PaperRank among authors could be proportional to the order of authors, rather than being uniform, as sketched in Remark 1.

Furthermore, to obtain indices that also take into account both the number of papers and their quality, it could be possible to reformulate h-index, h^c-index, and i10-index by replacing, in the original definition, the number of citations of each paper with the ratio between PaperRank and the number of authors. For example, we could define the following new indices, for an author \( A_k \), whose AuthorRank is given by (5), which correspond to the
h-index and i10-index, respectively:

- \( h_\alpha \)-index: the largest value \( p \) s.t. \( \sum_{j=1}^{m_k} \left( \frac{\text{PaperRank}(P_{b,j,k})}{\#\text{Auth}(P_{b,j,k})} \geq \alpha p \right) \geq p \),

- \( i_\beta \)-index = \( \sum_{j=1}^{m_k} \left( \frac{\text{PaperRank}(P_{b,j,k})}{\#\text{Auth}(P_{b,j,k})} \geq \beta \right) \),

where \( \alpha \) and \( \beta \) are positive quantities suitably chosen. In Table 4 we list some examples for such indices, for the authors with h-index=48 listed in Table 1 by choosing \( \alpha = 0.01 \) and \( \beta = 0.1 \), and compare them with the usual h-index and i20 index, respectively. We observe that the “outlier” author not shown in Figs. 1 and 2 (A.Id 7202686127) exhibits values similar to most of the

\[2\] Indeed, we found the latter index more significant than i10.
other authors. Moreover, the difference between the two authors in Fig. 2 in the “Physics and Astronomy” subject area is still very evident. This means that the combination of AuthorRank and one of these new parameters may provide a more comprehensive information on a given author.

Obviously, by considering only the most recent papers, one may derive an analog of the h*-index (see also [1]).

Last but not least, it is worth mentioning that one could, in principle, evaluate journals or research groups (or Universities) by combining the associated PaperRank and AuthorRank: this could represent a subsequent goal of this research.

6 Conclusions

In this paper, we show the result of a preliminary study on the application of the PaperRank and AuthorRank indices proposed in [1] to the Scopus database. The two indices give more interesting values than the bare number of citations and h-index, respectively, with practically the same computational cost. This fact allows to easily keep them always updated. They can also be used for the definition of new indices and to evaluate, always quantitatively, scientific journals and research groups.
Acknowledgement

We are very grateful to Elsevier, in particular to the Integration Support of Scopus, for the assistance and for granting us the intensive use of the API keys needed for this research. We are also grateful to the reviewers, for their comments and remarks.

References


[28] *Google* web page: [http://www.google.it/](http://www.google.it/)


[31] [https://dev.elsevier.com/sc_apis.html](https://dev.elsevier.com/sc_apis.html)