A simple centrality index for scientific social recognition

Osame Kinouchi *, Leonardo D.H. Soares, George C. Cardoso

Departamento de Física, FFCLRP, Universidade de São Paulo, Ribeirão Preto, 14040-901, Brazil

HIGHLIGHTS

- We discuss centrality indexes in the bipartite scientific papers–authors networks.
- We propose a new index (the K-index), which is simple to calculate and outperforms the Hirsch h-index in several aspects.
- The K-index correlates better with Nobel prizes than any of the other indexes furnished by the *Web of Science* (*WoS*) platform.
- The K-index is related to connections to the hubs of the scientific citation and social networks.
- The outliers in the K vs h plane contain interesting information that can be used for scientific prizes prediction.

ARTICLE INFO

ABSTRACT

We introduce a new centrality index for bipartite networks of papers and authors that we call K-index. The K-index grows with the citation performance of the papers that cite a given researcher and can be seen as a measure of scientific social recognition. Indeed, the K-index measures the number of hubs, defined in a self-consistent way in the bipartite network, that cites a given author. We show that the K-index can be computed by simple inspection of the Web of Science platform and presents several advantages over other centrality indexes, in particular Hirsch h-index. The K-index is robust to self-citations, is not limited by the total number of papers published by a researcher as occurs for the h-index and can distinguish in a consistent way researchers that have the same h-index but very different scientific social recognition. The K-index easily detects a known case of a researcher with inflated number of papers, citations and h-index due to scientific misconduct. Finally, we show that, in a sample of twenty-eight physics Nobel laureates and twenty-eight highly cited non-Nobel-laureate physicists, the K-index correlates better to the achievement of the prize than the number of papers, citations, citations per paper, citing articles or the h-index. Clustering researchers in a K versus h plot reveals interesting outliers that suggest that these two indexes can present complementary independent information.

1. Introduction

Complex networks is an interdisciplinary area where statistical physicists have done impressive contribution [1,2]. Due to an interest in complex networks with large and reliable data banks, physicists have frequently worked in Scientometrics, say in citation and scientific collaboration networks [3–9]. Actually, one of the founding fathers of Scientometrics is the physicist Derek J. de Solla Price. A contribution that increased very much the interest of physicists on Scientometrics was the

* Corresponding author.
E-mail address: osame@ffclrp.usp.br (O. Kinouchi).

introduction of the *h*-index, also due to a physicist, Jorge E. Hirsch [10]. Research in centrality indexes for citation networks experienced a boom after 2005 [11–15].

One of the decisive advantages of the *h*-index over its competitors is the ease of calculation even before platforms such as Web of Science (*WoS*) and Google Scholar implemented it as an automatic feature. The centrality *h*-index has been used not only to evaluate individual researchers but also single papers [16,17], scientific journals [18,19], universities [20] and countries [21].

Despite its popularity, the *h*-index has some well-known drawbacks [15,22–24]. An important one is that high impact scientists with a relatively small number of papers N cannot have a high h since $h \leq N$. Another issue is that scientists who clearly have different scientific recognition might have very similar *h*-indexes. One of the reasons for this ambiguity is that the *h*-index has a narrow dynamic range: rankings use integers from 0 to about 120 [10,25].

The *h*-index is also prone to distortions due to self-citations [22,26,27]. There are also issues related to dependence on the scientific area, although this can solved by proper normalization [11,28,29]. Lastly, it is not clear how the *h*-index correlates with more qualitative indicators of impact such as scientific prizes [30,31].

Here we introduce a new index, which we call the *K*-index, designed to address the issues discussed above. The *K*-index does not depend on N, has large dynamic range and is very robust to self-citations. In addition, the *K*-index correlates better to scientific prizes than the number of papers N, the number of citations C, the average number of citations per paper C/N, the number of citing articles CA and the *h*-index, that are the indexes readily available on *WoS*.

Furthermore, in contrast to the vast majority of the bibliometric indexes discussed in the literature [15], the *K*-index can be easily determined by simple inspection of the *WoS* platform. Its calculation could potentially be automated on the *WoS* in the future, in a manner similar to the calculation of the *h*-index.

The *K*-index is a new node centrality index, related to but different from the Lobby [32,33] and the $h(1)$ centrality indexes [34,35]. In addition, the *K*-index is competitive with the *h*-index, outperforming it in several important aspects.

2. Material and methods

2.1. The bipartite authors–papers complex network

First, we define the operator Θ that, when applied to a weighted network $\bar{W}_{ij}$, generates a binary network $W_{ij} \in [0, 1]$ following the Heaviside function $\Theta(\bar{W}_{ij}) = W_{ij}$.

Scientometric networks are not simple, but composed of several different layers and intermingled networks. For example, by only considering the open access data furnished by the *WoS* platform, we obtain several complex networks:

- The author collaboration network $\bar{A}_{ij}$: this is an undirected weighted network where nodes (authors) a_i and a_j are linked by the (integer) number $\bar{A}_{ij}$ of co-authored papers;
- The social collaboration network A_{ij}: the adjacency binary network $A_{ij} = \Theta(\bar{A}_{ij})$;
- The publication network P_{ik}: the undirected binary matrix $P_{ik} = 1$ if author node a_i published paper node p_k and zero otherwise;
- The citation network $\bar{C}_{kl}$: the directed weighted network where node paper p_l cites (an integer number of times) $\bar{C}_{kl}$ the same paper p_k;
- The binary citation network C_{kl}: the directed binary network $C_{kl} = \Theta(\bar{C}_{kl})$;
- The self-citation network S_{ikl}: a directed binary matrix where paper p_l cites the paper p_k with the constraint that both are authored by a_i. This can be written as $S_{ikl} = P_{ik}P_{il}C_{kl}$. We observe that the self-citation network S_{ikl} sometimes is blamed as a factor that inflates bibliometric indexes. This is a polemic topic, since a moderate level of self-citation can be a legitimate scientific practice.
- The citing articles network C_{il}^A: the directed binary network C_{il}^A where paper p_l cites at least some paper authored by a_i. Notice that if paper p_l cites two or more papers published by a_i, this counts as a single citing paper, that is, $C_{il}^A = 1$. This network matrix can be written as $C_{il}^A = \Theta\left(\sum_k P_{ik}C_{kl}\right)$.

2.2. Centrality indexes in the WoS

We will use the following notation for the centrality indexes provided by *WoS* for the *i*th researcher :

- Number of papers $N_i = \sum_k P_{ik}$;
- Number of citations $C_i = \sum_{k,l} P_{ik}C_{kl}$;
- Number of citations without self-citations $C_i' = C_i - \sum_{k,l} S_{ikl} = \sum_{k,l} P_{ik}(1 - P_{il})C_{kl}$;
- Number of citations per paper $C_i/N_i = \sum_{k,l} P_{ik}C_{kl} / \sum_k P_{ik}$;
- Number of articles that cite the researcher (or citing articles) $CA_i = \sum_l C_{il}^A$;
- Citing articles without self-citations $CA_i' = \sum_l (1 - P_{il})C_{il}^A$;
- Hirsch index h;
- Hirsch index h' without self-citations.

2.3. Calculation of the K-index

If a researcher has K citing papers, each one with at least K citations, then their K-index is K. By construction, the K-index measures not the raw quantity but the quality and importance of the citing articles: only highly cited papers that have cited the researcher enter in the calculation.

In contrast to most scientometric centrality indexes [15], the K-index is easily calculated by simple inspection of the *WoS* platform. We assume that other platforms like Google Scholar Citations or Harzing Publish-Perish could be easily adapted to provide K without difficulty. On the *WoS*, the procedure to get the K of a researcher is the following:

- Search the papers of a given author;
- Click on the link *Create Citation Report*;
- Click on the *Citing Articles* (CA) link (or Citing Articles without self-citations (CA'), if desired);
- Have the list of citing articles ranked from the most cited (defined as $r = 1$) to the least cited paper (this is the default ranking presented by *WoS*);
- Compare the citations $c(r)$ each citing article received to its rank r. In a descending order we have $r < c(r)$ up to a number K where $K = r \leq c(r)$ but $K + 1 = r + 1 > c(r)$. This defines the K-index.

The procedure is very similar to the calculation of the h-index, but now we use information from the second layer of citations, i.e., citations to the citing papers CA.

2.4. Correlation with other scientometric indexes

To study the correlation between the K-index and indexes N, C, C/N, CA and h, we have sampled 28 recent Nobel-laureate physicists and 28 highly cited (but not laureate) physicists, the latter sample randomly chosen from the Thompson-Reuters page Highly cited Physics researchers 2014 (http://highlycited.com). For each researcher, we obtained N, C, C/N, CA, h and K from *WoS*. Physicists from the past such as A. Einstein, P. Dirac and E. Ising, and also 2016 Nobel prize winners, were used for some comparisons but not for the study of correlations between indexes.

While N is the raw productivity of a researcher and C, C/N and CA are generally considered as measures of impact, the h-index is viewed as a combination of productivity and impact. We propose that the K-index is associated to social recognition by scientific hubs. The K-index reflects citations by highly cited authors which probably are the hubs of the social scientific network A_{ij} (or at least the authors of the paper hubs of the citation network C_{kl}).

3. Results

First, we examine the correlation between the new K-index and the standard indexes extracted from *WoS* for our two groups of scientists. Second, we study the correlation between the ranking produced by several indexes and Nobel prizes received by the sampled researchers.

3.1. Comparison to standard indexes

In Fig. 1, we compare the K-index with the standard indexes furnished by *WoS*. The correlations between K and N (Fig. 1a) and K and C/N are weak (Fig. 1c), indicating that the K-index can give non-redundant information. There is a noticeable correlation between K and CA (Fig. 1d), but the same correlation happens for h vs CA (not shown), so this is not a particular weakness of the K index. Indeed, the correlation of Fig. 1d is somewhat misleading. For example, the K-index is much more robust to self-citations than CA. When self-citations are considerable, CA's growth could follow an arithmetic progression – of order N^2 – while K-index depends only on the K-core of highly cited articles. Also, since CA counts poorly cited, perhaps irrelevant, papers, it cannot detect fraudulent behavior, as we will discuss later in Subsection 4.6, where the researcher has high $CA = 1051$ but very low $K = 46$.

The correlation plots of Fig. 1 do not distinguish in a clear way between Nobel laureates and not laureates, that is, all researchers basically form a single cluster. In contrast, the K vs h plane brings unique information and suggests that laureates and non-laureates form different clusters, as we will discuss now.

3.2. The K versus h plane

The K and h indexes seem to convey complementary information, as shown in Fig. 2. Despite the overall correlation (Pearson correlation $= 0.7$), the interesting features are the outliers. We suggest the following classification, where the words true and false refer to the expectation created by the h ranking:

- True positives (high h, high K): researcher with both high production, impact and recognition.
- True negatives (low h, low K): researchers with low important production, impact and impact.
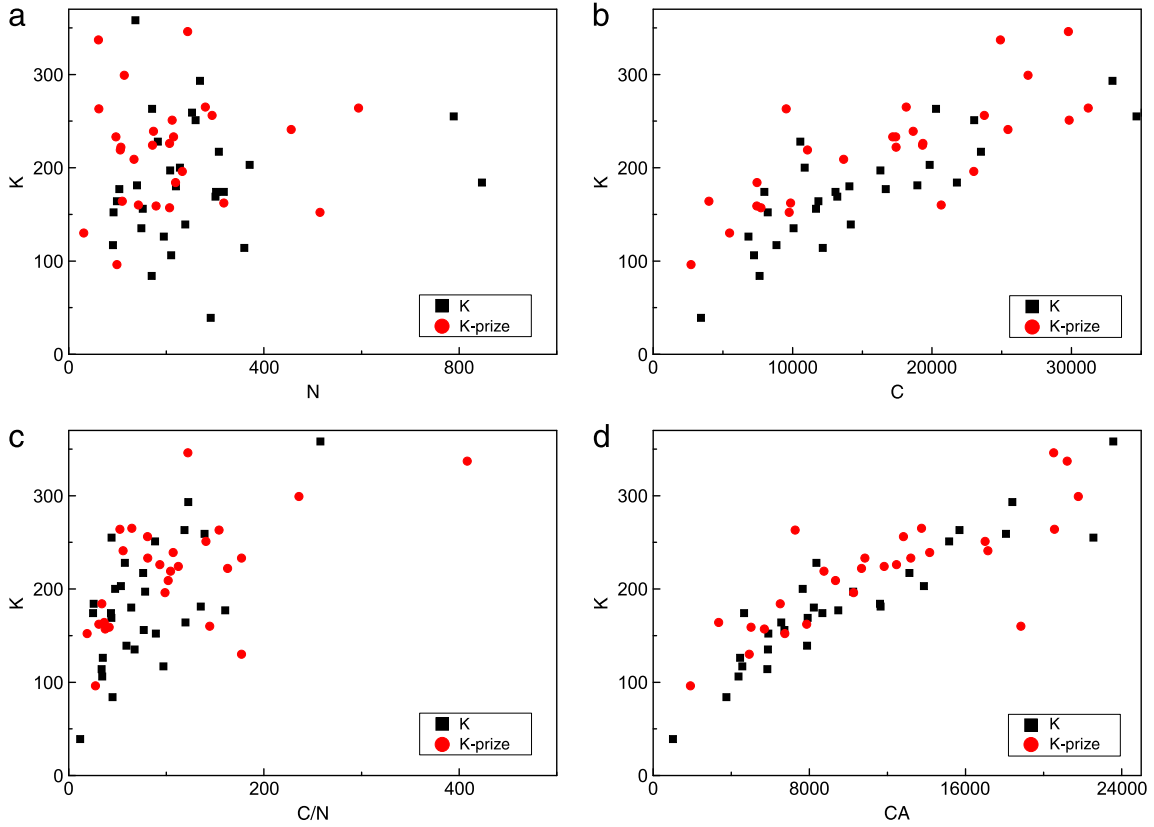- False negatives (low h, high K): researchers with few papers with high scientific recognition.

Fig. 1. Comparison between the *K*-index and standard *WoS* bibliometric indexes: (a) *K* versus *N*; (b) *K* versus *C*; (c) *K* versus *C/N* and (d) *K* versus *CA*. The squares represent researchers who did not received Nobel prizes, while the circles represent laureates.
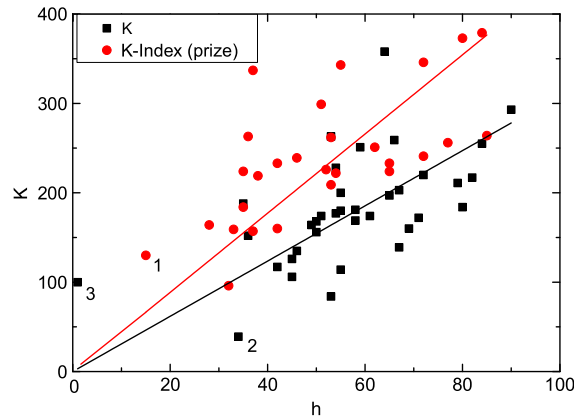

Fig. 2. *K*-index vs *h*-index for Nobelists (circles) and non-Nobelists (squares). 1: Paul Dirac; 2: Mohamed El Naschie (scientist with controversial case of self citations); 3: Ernst Ising; The two lines are linear fitting curves for the Nobel Prizes and the non-laureate highly cited physicists that suggest the presence of two different clusters.

- False positives (high *h* low *K*): researchers with possibly inflated *h* or with large production *N* and high citation *C*, but with no exceptional scientific recognition.

As suggested by the different inclination of the curves in Fig. 2, for a similar *h*, Nobelists tend to have a higher *K*. However, there are four main types of anomalies:

- Nobel laureates with low *K* and *h*. This anomaly seems to be due to the area of research (say, a low average number of papers *N*) and could be fixed by proper normalization [11,28,29].
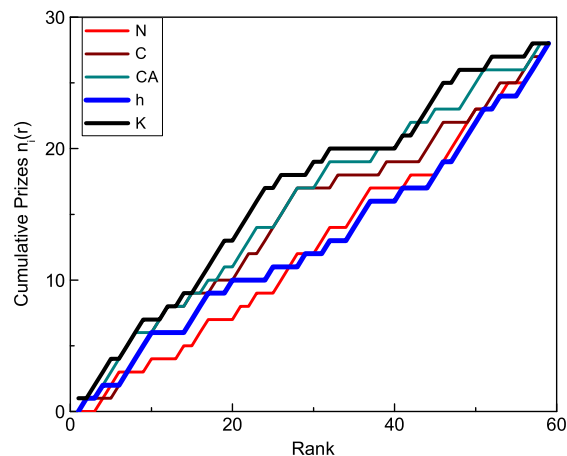
Fig. 3. Cumulative curve $n(r)$. Number of researchers $n(r)$ with Nobel prizes as a function of their relative rank r for each index classification listed in the inset. The higher the value $n(r)$ for a given r, the better the correlation between the index and Nobel prizes.

- Non-laureates with very high K and h. Here the anomaly could be either because Nobel prizes have some known biases (e.g., less prizes for pure mathematical physicists, despite high citations) or refer to non-laureated researchers with high prospects to be laureated in the near future.
- False negative (in terms of h): the case of Ernst Ising (point 3 in Fig. 2, see Subsection 4.2), who has $h = 1$ from a paper authored solo [36]. Despite such a low h, K-index = 100 characterizes Ising as a pioneer on a fertile topic, achieving social recognition by the hubs of the scientific social network.
- False positive (in terms of h): the controversial case of El Nashie (point 2 in Fig. 2, see Subsection 4.6) illustrates the case of a false positive: $h = 35$ but $K = 46$ only.

3.3. Nobel prize correlation curve

We constructed a correlation curve between indexes and Physics Nobel prizes as follows: for each index (N, C, CA, h, K), we ranked the researchers of our sample. In the horizontal axis, we placed scientists from highest ($r = 1$) to lowest ($r = 56$) ranks. For each index, we build a ranked list of the researchers, from highest to lowest. Then, descending along the list, we compute the number $n(r)$ of Nobel Prizes found in researchers ranked up to rank r. The faster $n(r)$ grows the more sensitive and specific is the index for the identification of laureates. The curve $n(r)$ measures how many Nobel Prizes (n) occur up to the rank r as defined by each index.

The larger the area below the curve $n(r)$, the better the performance of the corresponding index. Fig. 3 contains one of the main proofs of superiority of K-index in scientific social recognition identification: if we rank researchers by the indexes N, C, CA, h and K, we observe that the curve $n(r)$ for the K-index gives the best sensitivity and correlation with Nobel prizes. Interestingly, K-index performs well above the h index.

4. Discussion: Why is the K-index superior to standard centrality indexes?

4.1. The K-index is very easy to compute

An extensive and comprehensive compilation of bibliometric indexes can be found in a book by Todeschini and Baccini [15]. It is natural to ask if there is any gain in introducing yet another index such as the K-index. The idea of using the second layer of citations has been advocated by Hu et al. [37], although not in the Hirsh-like form of the K-index. Indeed, one could argue that the K-index is related to the complex networks Lobby index [32,33,38]. For this, we need to change the Lobby index original definition so that, instead of papers being single nodes, we collapse the total set of the researcher papers into a large and unique macro-node, and the citing articles would be considered as the other nodes in a (directed) network.

Our claim is not that there are no similar ideas in the literature. Our main claim is that the K-index, when compared to similar indexes, has an important and unique advantage: it is very easy to compute in the *WoS* platform. Any researcher, by using *WoS*, can calculate their K-index. The same cannot be said about most of the indexes, as can be seen in the aforementioned handbook [15]. We presume that, if the value of the K-index is recognized, *WoS*, *Google Citations* and other data banks can easily implement an automatic calculation of the K-index, as has already been done for the h-index.

4.2. The K-index does not depend on N

We believe that one great advantage of the K-index is its indirect dependence on the number N of papers published by the researcher. K has no upper bound while $h \leq N$. An extreme example is Ernst Ising, who published an influential model [36], derived from his Ph.D. thesis, in a single author paper. After that, due to Nazi persecution, Ising emigrated from Germany to Luxembourg and stopped to do any research. In 1947, he emigrated to the United States and becomes a professor at Bradley University. His 1925 paper received 1373 citations at the time of this writing. The only other paper Ising would publish, indeed a commentary at the *American Journal of Physics* about Goethe's optics [39], has a single citation in *WoS*. Therefore, Ising h-index is equal to one (by the way, with our present citation, Ising h-index will grow to its maximum value $h = 2$). It seems unfair to give such a low ranking to someone who published the pioneering idea on what is now known as the Ising model. The h-index puts Ising in the same ranks of an undergraduate student with a single paper cited a single time. The K-index tells a very different story: Ising has a considerable value $K = 100$. This is fairer and puts Ising well above students with single papers. The K-index makes justice to Ising, indicating his scientific social recognition. Another example is Paul Dirac, who has a low $h = 15$. However, he has a $K = 130$, signaling also social recognition by other scientists (indeed, there are four Physics Medals in honor of Dirac). This low value for h is not due to the fact that Dirac is a scientist from the beginning of the 20th century, but seems to be correlated to his lower productivity N, which is a constraint for his h-index.

4.3. The K-index has larger classification range than h-index

As observed in the K vs h plane of Fig. 2, the range of K is at least a factor of three times larger than the range of h. This implies that K produces a finer discrimination between researchers. This is a clear advantage for ranking purposes. On the other hand, the K-index continues to be based on integers. This feature gives more stability to ranking since real numbers ranks are more prone to spurious ranking order due to small statistical fluctuations.

4.4. K-index better discriminates scientists with the same h-index: The Einstein–Hirsch paradox

The h-index frequently presents paradoxical results on its ranking, giving similarly large h to researchers with very different scientific recognitions. For an example, Jorge Hirsch's h-index is 55, while Albert Einstein's is 51. It seems that, due to its limited classification range, the h-index has poor discrimination power for researchers with large N and C. The K-index for Hirsch is 200, which is high, but Einstein's is superior: $K = 299$. This is another demonstration of how K-index can be a powerful tool to discriminate between scientists of very different scientific social recognition and similar h. Let us take a look at another example: Einstein ($h = 51$; $K = 299$) and Edward Witten, the most cited physicist of the world ($h = 120$; $K = 368$). While Witten's h is 2.35 times Einstein's h, Witten's K is only 1.23 times Einstein's h, what seems to be more fair in terms of scientific social recognition.

4.5. The K-index is very robust to self-citations

Several studies have shown that the h-index is prone to manipulation by self-citations [26,27,40]. It is very easy to see why the K-index is robust to self-citations. Only highly cited papers compose the K-core. If a researcher makes a self-citation, it will only be counted for K if it is of an already highly cited paper. Of course, some difference $\Delta K = K - K'$ may appear between the K-index calculated from all citing articles and the K'-index due to citing articles without self-citations (also furnished by *WoS*). Our claim is that $\Delta K/K$ is much smaller than the corresponding $\Delta h/h$. *En passant*, we notice that the K'-index (K calculated without self-citations) is easily obtainable from *WoS* while the h'-index (h calculated without self-citations) is not furnished or directly calculable in the platform.

4.6. The K-index detects scientific career frauds

A public instance of scientific fraud has been denounced by *Nature* [41]. The researcher, El Naschie, who was the editor of a scientific journal, was suspected of scientific misconduct. Superficially, his scientific career looked impressive: $N = 293$ and $h = 35$, which would put him among top ranking physicists. We could compare El Nashie with known physicists such as Robert H. Swendsen ($N = 149$; $h = 37$) and Ronald Dickman ($N = 134$; $h = 28$). Indeed, for El Naschie we have only $K = 46$, in contrast to $K = 206$ for Swendsen and $K = 76$ for Dickman. Also, the K-index calculated by using the *CA* without self-citations is $K = 39$ (El Naschie), $K = 204$ (Swendsen) and $K = 71$ (Dickman), showing that, proportionally, the K index is much more affected by self-citations in the abnormal case: $K/h = 1.31$, $\Delta = (K - K')/K' = 0.18$ (El Naschie), $K/h = 5.57$, $\Delta = 0.01$ (Swendsen), $K/h = 2.71$, $\Delta = 0.07$ (Dickman). We suggest that unusually low K/h and K/N and a higher Δ ratios could be indicative of abnormal publication patterns.

Table 1
Statistics for the sample of laureates for different indexes. The K-index has the lowest coefficient of variation (CV) among competing indexes.

	K (prize)	CA (prize)	h (prize)
Mean	224	12792	52
Standard deviation	66	8286	18
CV	29%	65%	35%

4.7. The K-index correlates well with scientific prizes

Some literature studies have discussed the relationship between centrality indexes, specifically the h-index, and scientific success in terms of future performance and recognition such as scientific prizes [13,42–44]. The K-index correlates better with Nobel prizes than N, C, C/N, CA and h, as shown in Fig. 3. Indeed, it is surprising that the h-index performs worse for this task than any of the other indexes compared. In addition, the K-index has the smallest coefficient of variation (CV) for Nobelists, see Table 1. This means that the K-index correlation with Nobel Prizes is tighter than that of the other indexes.

Here we must make a disclaimer: correlation is not prediction. We cannot to say that a high enough K-index predicts that the researcher will win a prize. Our test sample included researchers who received the prize in the last 18 years, most of them in the last 10 years; probably C and CA increase after the prize—the pattern was not clear from the data. To study K as a predictor of scientific prizes we must count only the citations received by a laureate scientist before the prize (or study scientists that have not yet been awarded a prize and follow up for possible prizes). An illustrative example is given by the 2016 Physics Nobel Prizes. By using their metrics, *WoS* made a list of probable winners a few days before the announcement of the prize. It seems that *WoS* metrics correlates well with the h-index, since all selected researchers have a very large h: J.A. York ($h = 79$, $K = 211$), R.K.S. Thorn ($h = 72$, $K = 220$), C. Grebogi ($h = 69$, $K = 160$) and W. Drever ($h = 50$, $K = 168$). However, the true winners have a lower h with a higher K (all above $K = 200$): D.J. Thouless ($h = 55$, $K = 343$), F.D.M. Haldane ($h = 53$, $K = 262$) and J.M. Kosterlitz ($h = 35$, $K = 224$). This anecdotal evidence suggests that ranking by using K seems to better predict the Nobel Prize than ranking by h. Of course, this must be checked by using larger samples. Our point is that for the 2016 Nobel Prize, the K-index has passed the test when compared with the *WoS* prediction: the actual winners had lower h than most of the researchers predicted by *WoS*, while all had higher K than those of the *WoS* selection.

4.8. Limitations of the K-index

Since K does not depends on N, it is a scientific impact index, not a productivity index. As we have seen, Ernst Ising with $N = 2$ has a large K-index of 100. But someone with few papers that publishes (possibly with seniors researchers) a well-cited review or participates in large science collaborations might have a large K and a small h. These outliers with small h and high K in the K vs h plane must be individually analyzed. Some are big science low N collaborators, some are authors of reviews, but others are low N high impact scientist, even Nobel prize winners (Dirac). Inflation of K due to publishing reviews can be mitigated by eliminating the reviews in the researcher's *WoS* statistical summary before searching for citing articles CA, but the other outlier instances must be examined case by case.

The K-index does not discriminate the following (extreme) situation. Consider two researchers, say Alice with $h = 30$ and $K = 80$, where her K is due to citations for all her papers. The other researcher, say Bob, also has $h = 30$ and $K = 80$, but now his K is due to citations to a single important paper. Although artificial and rare, this extreme case could occur and K does not discriminate Alice from Bob. However, the h-index cannot distinguish between them either.

This particular example only confirms a trivial fact in Scientometrics: to fully characterize the bibliographic production of a researcher, instead of a small set of scalar indexes, we need the full networks P_{ik}, $\bar{C}_{kl}$, A_{ij} of the bipartite network of scientific papers and authors. And, if we want to compare researchers from different scientific areas, some kind of normalization (in N, C, C/N, CA, h, K etc.) must be done so that the distributions collapse into universal curves as shown in [11,28,29].

4.9. Comparison with other new indexes

In this paper we compare our centrality K-index only with the standard indexes N, C, C/N, CA and h. No comparison is made with new indexes such as the g index [12] and the h_I individual Hirsch index [11]. We chose to work with the *WoS* database, where the K-index can be easily determined, while the computation of indexes such as g index and h_I is not trivial. We defer a comparison between the K-index and such indexes to another paper.

It is important to note that, although inspired by, the K-index is not equivalent to the Lobby [32,33] or the $h(1)$ [34,35] indexes applied to citation networks. The $h(1)$ index, as a natural extension of the h-index (called $h(0)$), would be determined as follows: a researcher has index $h(1)$ if he/she has $h(1)$ papers with $h(1)$ citations, each one with at least $h(1)$ citations. This index is also bounded, $h(1) \leq N$. This is very different from the way the K-index is defined. In addition, $h(1)$ cannot be readily determined by quick inspection in the *WoS* platform as the K-index can.

Out of 150 bibliometric indexes listed in [15], none of them can be easily calculated by inspection the *WoS*, with the exception of the h-index. This fact alone makes the K-index superior, in the sense of easy of determination, to all those centrality indexes. And easy calculation is a pivotal property for a centrality index to become accepted and popular.

4.10. Generalizations of the K-index

As has been done to other indexes, we can generalize and adapt the K-index to evaluate research impact in a number of contexts:

- **Proximal K_m index:** in the calculation of the K-index, only articles from the researcher published in the last m years are considered. This can be easily done in the *WoS* by selecting the time interval of the search.
- **Recent impact K_y index:** in the calculation of the K-index, only citing articles published in the last y years are accounted. This can be done in the *WoS* by sorting the citing articles by year, eliminating those published before the last y years, then sorting anew by citation importance and calculating the K-index as before.
- **Group K_g index:** As in the case of an individual researcher, we can compute the citing articles received by a group (research team, department, university, country etc.). For department, university and country it is currently possible for open access users of the *WoS* to compute the K-index if the number of records is lower than 10,000. We can work around this limitation if we select the search for a limited number of years for departments and universities, and for single years (or semesters) for countries. So, we can also obtain a $K_g(t)$ index that is a function of time.

5. Conclusion and perspectives

We have introduced a new citation index (the K-index) that, like the h-index, is easily obtainable from the *WoS* platform. The K-index outperforms the Hirsch index in several aspects: has a larger classification range, distinguishes researchers with similar h, does not depend on the number N of papers, is robust to self-citations, is a good detector of cases of some scientific frauds and has better correlation to scientific prizes. The plane K vs h also gives interesting information, as exemplified by the data for the 2016 Nobel Prize. The study of this plane as a predictor of scientific prizes seems to be very promising, by using other samples such as the Wolf prize, the Boltzmann medal, and Maxwell medal.

Acknowledgments

References

[1] M. Newman, Networks: An Introduction, Oxford University Press Inc., United Slates, New York, 2010, pp. 1–2.
[2] A.L. Barabási, Network Science, Cambridge University Press, 2016.
[3] M.E. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. 98 (2) (2001) 404–409.
[4] A.L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A 311 (3) (2002) 590–614.
[5] M.E. Newman, Coauthorship networks and patterns of scientific collaboration, Proc. Natl. Acad. Sci. 101 (suppl 1) (2004) 5200–5205.
[6] M. Wang, G. Yu, D. Yu, Measuring the preferential attachment mechanism in citation networks, Physica A 387 (18) (2008) 4692–4698.
[7] F.X. Ren, H.W. Shen, X.Q. Cheng, Modeling the clustering in citation networks, Physica A 391 (12) (2012) 3533–3539.
[8] J.R. Clough, T.S. Evans, What is the dimension of citation space? Physica A 448 (2016) 235–247.
[9] Z. Xie, Z. Ouyang, Q. Liu, J. Li, A geometric graph model for citation networks of exponentially growing scientific papers, Physica A 456 (2016) 167–175.
[10] J.E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA (2005) 16569–16572.
[11] P.D. Batista, M.G. Campiteli, O. Kinouchi, Is it possible to compare researchers with different scientific interests? Scientometrics 68 (1) (2006) 179–189.
[12] L. Egghe, An improvement of the *h*-index: The *g*-index, ISSI Newsl. 2 (1) (2006) 8–9.
[13] J.E. Hirsch, Does the *h*-index have predictive power? Proc. Natl. Acad. Sci. 104 (49) (2007) 19193–19198.
[14] M. Schreiber, Twenty Hirsch index variants and other indicators giving more or less preference to highly cited papers, Ann. Phys. 522 (8) (2010) 536–554.
[15] R. Todeschini, A. Baccini, Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research, John Wiley & Sons, 2016.
[16] A. Schubert, Using the *h*-index for assessing single publications, Scientometrics 78 (3) (2008) 559–565.
[17] M. Martínez, M. Herrera, J. López-Gijón, E. Herrera-Viedma, H-classics: Characterizing the concept of citation classics through *h*-index, Scientometrics 98 (3) (2014) 1971–1983.
[18] T. Braun, W. Glänzel, A. Schubert, A Hirsch-type index for journals, Scientometrics 69 (1) (2006) 169–173.
[19] C. Malesios, Measuring the robustness of the journal *h*-index with respect to publication and citation values: A Bayesian sensitivity analysis, J. Informetrics 10 (3) (2016) 719–731.
[20] G. Abramo, C.A. DAngelo, F. Viel, The suitability of *h* and *g* indexes for measuring the research performance of institutions, Scientometrics 97 (3) (2013) 555–570.
[21] E. Csajbók, A. Berhidi, L. Vasas, A. Schubert, Hirsch-index for countries based on essential science indicators data, Scientometrics 73 (1) (2007) 91–117.
[22] B. de Keijzer, K.R. Apt, The *h*-index can be easily manipulated, Bull. EATCS 2 (2013) 79–85.
[23] L. Bornmann, R. Mutz, H.D. Daniel, Are there better indices for evaluation purposes than the *h*-index? A comparison of nine different variants of the *h*-index using data from biomedicine, J. Amer. Soc. Inf. Sci. Technol. 59 (5) (2008) 830–837.
[24] L. Waltman, N.J. Van Eck, The inconsistency of the *h*-index, J. Amer. Soc. Inf. Sci. Technol. 63 (2) (2012) 406–415.
[25] M.A. García-Pérez, An extension of the *h*-index that covers the tail and the top of the citation curve and allows ranking researchers with similar *h*, J. Informetrics 6 (4) (2012) 689–699.
[26] M.H. Huang, W.Y.C. Lin, Probing the effect of author self-citations on *h*-index: A case study of environmental engineering, J. Inf. Sci. 37 (5) (2011) 453–461.

[27] M. Schreiber, Self-citation corrections for the Hirsch index, Europhys. Lett. 78 (3) (2007) 30002.
[28] J.E. Iglesias, C. Pecharromán, Scaling the *h*-index for different scientific ISI fields, Scientometrics 73 (3) (2007) 303–320.
[29] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, *h*-index: A review focused in its variants, computation and standardization for different scientific fields, J. Informetrics 3 (4) (2009) 273–289.
[30] A. Mazloumian, Y.H. Eom, D. Helbing, S. Lozano, S. Fortunato, How citation boosts promote scientific paradigm shifts and Nobel prizes, PLoS One 6 (5) (2011) e18975.
[31] G. Abramo, T. Cicero, C.A. DAngelo, Are the authors of highly cited articles also the most productive ones? J. Informetrics 8 (1) (2014) 89–97.
[32] A. Korn, A. Schubert, A. Telcs, Lobby index in networks, Physica A 388 (11) (2009) 2221–2226.
[33] M.G. Campiteli, A.J. Holanda, L.D. Soares, P.R. Soles, O. Kinouchi, Lobby index as a network centrality measure, Physica A 392 (21) (2013) 5511–5515.
[34] L. Lü, T. Zhou, Q.M. Zhang, H.E. Stanley, The *h*-index of a network node and its relation to degree and coreness, Nat. Commun. 7 (2016).
[35] R. Pastor-Satorras, C. Castellano, Topological structure and the *h*-index in complex networks, Phys. Rev. E 95 (2) (2017) 022301.
[36] E. Ising, Beitrag zur theorie des ferromagnetismus, Z. Phys. A 31 (1) (1925) 253–258.
[37] X. Hu, R. Rousseau, J. Chen, On the definition of forward and backward citation generations, J. Informetrics 5 (1) (2011) 27–36.
[38] R. Rousseau, S.X. Zhao, A general conceptual framework for characterizing the ego in a network, J. Informetrics 9 (1) (2015) 145–149.
[39] E. Ising, Goethe as a physicist, Amer. J. Phys. 18 (4) (1950) 235–236.
[40] G.A. Vîiu, A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation, J. Informetrics 10 (2) (2016) 552–566.
[41] Q. Schiermeier, Self-publishing editor set to retire, Nature 456 (7221) (2008) 432–433.
[42] P. Vinkler, Eminence of scientists in the light of the *h*-index and other scientometric indicators, J. Inform. Sci. 33 (4) (2007) 481–491.
[43] J. Hönekopp, J. Khan, Future publication success in science is better predicted by traditional measures than by the *h*-index, Scientometrics 90 (3) (2012) 843–853.
[44] O. Penner, R.K. Pan, A.M. Petersen, K. Kaski, S. Fortunato, On the predictability of future impact in science, Sci. Rep. 3 (2013).