Citation network centrality: a scientific awards predictor?

Osame Kinouchi*, Adriano J. Holanda and George C. Cardoso

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

Abstract

The K-index is an easily computable centrality index in complex networks, such as a scientific citations network. A researcher has a K-index equal to K if he/she is cited by K articles that have at least K citations. The K-index has several advantages over Hirsh's h-index and, in previous studies, has shown better correlation with Nobel prizes than any other index given by the *Web of Science*, including the h-index. However, correlation is not causation. Here we perform an experiment using the K-index, producing a shortlist to predict future Physics Nobel candidates. Our list has been built before the 2019 Nobel Prizes announcements and should be compared to the actual results of the current and following years.

Keywords: complex networks, node centrality, Hirsch index, Nobel prizes.

* Corresponding author.

Highlights

- We propose that the K-index could be a good network centrality index for the physics community and relevant to predict the likelihood of scientific prizes;
- We propose an experiment where a list of highly cited candidates is refined to predict Physics Nobel Prizes in the near future;
- We present a list with twelve candidates with highest K-index from an initial list of 138 physicists from *Clarivate Highly Cited Researchers 2019* (HCR).
- We present and discuss the K versus h plane for the candidates.

1. Introduction

Statistical physicists have made important contributions to the interdisciplinary area of complex networks [1, 2]. In particular, physicists have intensively studied scientometric networks thanks to the availability of large and reliable data banks [3, 4, 5, 6, 7, 8]. Indeed, an important advancement for the area came with the introduction of the h-index by physicist Jorge E. Hirsch [9]. A researcher has h-index h if he/she has published h papers each one with at least h citations. Centrality indexes proposals for citation networks experienced a boom after the introduction of the h-index [10, 11, 12, 13, 14].

A decisive advantage of the h-index over its competitors is its ease of calculation. However, it also is known that the h-index has several drawbacks. For example, if a researcher has published a small or moderate number N of papers, then necessarily $h \leq N$, even if every paper is of very high quality and has received thousands of citations.

Recently, we have proposed the K-index, centrality index that is complementary to h-index and also easy to calculate in the *Web of Science* (*WoS*) [15, 16]. In these publications, we verified that Physics and Physiology Nobel Prizes laureates have very high K-index (often above $K = 300$) but sometimes moderate h-index. However, this is only a correlation, and the growth of the K-index could have occurred after the acceptance of the prize.

Here we propose to test K-index's predictive power by using it to refine the *Clarivate Highly Cited Researchers 2019* (HCR) list of candidates to the 2019 Physics Nobel Prizes. Our task is a hard one since it depends only on a brute correlation between a scientometric index and the awards, and does not take into account nuanced and sociological guesses about the actual candidates.

2. Materials and Methods

2.1. The Highly Cited Researchers list of Clarivate Analytics

As a primary source, we used *Clarivate Highly Cited Researchers 2019* (HCR) list to furnish an initial sample of 138 candidates that have Research ID or Orcid. The methodology used by HCR to achieve this sample list is not of our concern now, and can be found in https://hcr.clarivate.com/methodology/purpose-and-methodology/. Our data and automated ranking script is available for public use at https://github.com/ajholanda/k-nobel/blob/master/k-nobel.md

Our task is to refine the HCR list by using the K-index. We will produce a shortlist of 12 candidates which is the maximum number of Nobel laureates for a period of four years and about 8.6% of the original HCR *2019* list.

For comparison, the K and h-indexes for the 138 physicists from the HCR list are presented in Fig. 1. The h-index is furnished directly by the *WoS*.

2.2. Calculation of the K-index

The K-index has been devised to measure the impact of the papers that cite a researcher, not just to measure the quantity or distribution of citations. If a maximum number of K papers cite a given author, each one with at least K citations, then the researcher has K-index equal to K [15, 16].

Centrality indexes that tried to improve the h-index, in general, involve impractical calculations [14]. The decisive advantage of the K-index is that it is easily determined by simple inspection of the *WoS* platform. We presume that other platforms like Google Scholar Citations could also be easily adapted to provide K automatically.

On the *WoS*, currently, one can obtain the K-index of a researcher from the following simple steps:

- Search the production of a given author;
- Click on the link *Create Citation Report*;
- Click on the link *Citing Articles* (*CA*) (or *Citing Articles without self-citations*, if desired);
- Have the list of citing articles ranked from the most cited (defined as rank $r = 1$) to the least cited (that is the default ranking presented by *WoS*);
- Compare the article rank r (on the left) with its citation count $c(r)$ on the right. When $r \leq c(r)$ but $r+1 > c(r+1)$, stop: the K-index is $K = r$.

3. Results and Discussion

In table I, we present twelve candidates from the *Clarivate* HCR list ranked by the K-index. In Fig. 1, we present the K versus h plane for all 138 researchers from the HCR, where the top 12 highest K-indexes are represented as red triangles. A filtered group, where graphene researchers have been removed, is represented by green inverted triangles. This has been done because graphene has already been the theme of a Nobel Prize in 2010.

Our objective with Fig. 1 is to show that scientists with high K do not necessarily have high h and vice-versa. K and h have complementary information. Fig. 1 should be compared to the K vs. h plots in [15, 16], where 28 Nobel laureates show K values well above other scientists' of similar h. However, it is not clear how much their K indexes have grown after receiving their prizes. A considerable inertial growth is a plausible hypothesis and correlation effects are difficult to separate.

Rank	Name	K	h
1	Michael Graetzel	611	206
2	Paul Alivisatos	605	145
3	Sergey V Morozov	559	37
4	Younan Xia	542	189
5	Pulickel Ajayan	541	147
6	Philip Kim	519	83
7	Zhong Lin (Z.L.) Wang	515	195
8	John P. Perdew	502	86
9	Yi Cui	495	124
10	Mikhail I. Katsnelson	489	89
11	Yang Yang	471	120
12	Alex K. Zettl	460	105

Table 1: List of our twelve Nobel Prize candidates as ranked by the K-index

Fig. 1 intends not to be correlational but predictive. All candidates have high citation rates and comparable h-index. However, we have chosen the top twelve K as our (crude) test of the predictive power of the K-index.

A limitation of our study is that our original sample is from *Clarivate* HCR; if possible nominee candidates are not initially on the list, then our present K ranking cannot detect and choose them.

Another limitation of our study is that the chosen names are, in principle, uncorrelated, that is, do not refer to a common discovery or a similar research topic. By contrast, the Nobel Prize for a given year is typically awarded to researchers who have made progress on similar topics or discoveries. Also, it is possible that researchers with very high K or h will not be laureated because they work on a topic that has already been awarded in the past.

Our methodology can be adjusted to incorporate the information discussed above. For example, we have noticed a very high proportion of authors with large K in the area of materials science, especially graphene research – which has already been an awarded topic in 2010. To correct for this bias, we have produced a second list where graphene researchers are removed. The new list with 12 candidates and with graphene scientists filtered out is given in Table 2 and has five new names as a replacement.

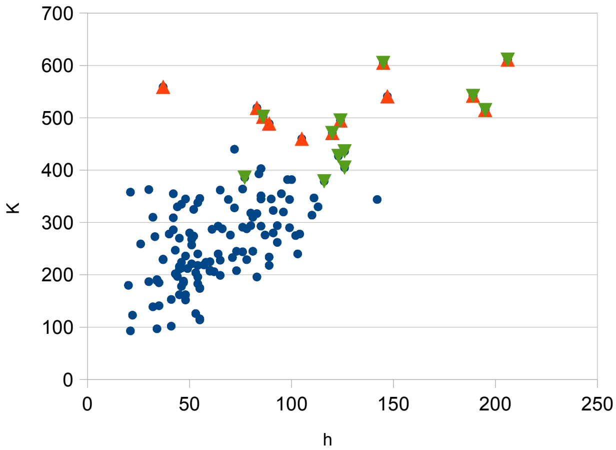

Figure 1: K-index vs h-index for the candidates of Table 1 (triangles), of Table 2 (inverted triangles) and other highly cited researchers (circles).

Rank	Name	K	h
1	Michael Graetzel	611	206
2	Paul Alivisatos	605	145
3	Younan Xia	542	189
4	Zhong Lin (Z.L.) Wang	515	195
5	John P. Perdew	502	86
6	Yi Cui	495	124
7	Yang Yang	471	120
8	Mohammad K Nazeeruddin	436	126
9	Naomi Halas	427	123
10	John Rogers	405	126
11	Arthur J. Nozik	386	77
12	Peter Zoller	379	116

Table 2: List of twelve Nobel Prize candidates as ranked by the K-index with graphene scientists filtered out.

4. Conclusion

It is an open question whether bibliometric information can have predictive power for scientific prizes. Prizes denote qualitative scientific recognition at the sociological level, where human factors are very important. Nobody would think that a prize should be decided by ranking the production of scientists by some automatic metric. At the same time, prizes intend to recognize original contributions whose impact is reflected in the bibliometric indexes, so it is plausible that predictive information is hidden in these indexes.

From a list of highly cited researchers, we proposed candidates for the 2019 or following years' Physics Nobel Prizes. We have presented a naive ranking and also an improved ranking where a citation bias for materials scientists studying graphene was filtered out. Our list of candidates can be updated and also used in future years.

The predictive study of this paper could be extended to other scientific prizes such as the Wolf Prize, the Boltzmann Medal, the Fundamental Physics Prize, and prizes in other scientific disciplines. The only difference is that the sample of initial candidates should be selected in accord with the specific scientific area. These predictive tests, perhaps in the form of annual contests, could be useful benchmarks for evaluation of centrality indexes that can then be used in other, less monitored and less well-studied, complex networks.

Acknowledgments

O. K. acknowledges financial support from the National Council for Scientific and Technological Development (CNPq) and CNAIPS – Center for Natural and Artificial Information Processing Systems (University of Sao Paulo). This paper results from research activity performed at The São Paulo Research Foundation (FAPESP) Center for Neuromathematics (FAPESP grant 2013/07699-0).

References

References

[1] M. Newman, Networks: an introduction, Oxford University Press, Inc., New York (2010) 1–2.
[2] A.-L. Barabási, Network Science, Cambridge University Press, Cambridge, 2016.
[3] M. E. Newman, The structure of scientific collaboration networks, Proceedings of the National Academy of Sciences 98 (2001) 404–409.
[4] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A: Statistical mechanics and its applications 311 (2002) 590–614.
[5] M. Wang, G. Yu, D. Yu, Measuring the preferential attachment mechanism in citation networks, Physica A: Statistical Mechanics and its Applications 387 (2008) 4692–4698.
[6] F.-X. Ren, H.-W. Shen, X.-Q. Cheng, Modeling the clustering in citation networks, Physica A: Statistical Mechanics and its Applications 391 (2012) 3533–3539.
[7] J. R. Clough, T. S. Evans, What is the dimension of citation space?, Physica A: Statistical Mechanics and its Applications 448 (2016) 235–247.
[8] Z. Xie, Z. Ouyang, Q. Liu, J. Li, A geometric graph model for citation networks of exponentially growing scientific papers, Physica A: Statistical Mechanics and its Applications 456 (2016) 167–175.
[9] J. E. Hirsch, An index to quantify an individual's scientific research output, Proceedings of the National academy of Sciences of the United States of America (2005) 16569–16572.
[10] P. D. Batista, M. G. Campiteli, O. Kinouchi, Is it possible to compare researchers with different scientific interests?, Scientometrics 68 (2006) 179–189.
[11] L. Egghe, An improvement of the h-index: The g-index, ISSI newsletter 2 (2006) 8–9.
[12] J. E. Hirsch, Does the h index have predictive power?, Proceedings of the National Academy of Sciences 104 (2007) 19193–19198.
[13] M. Schreiber, Twenty hirsch index variants and other indicators giving more or less preference to highly cited papers, Annalen der Physik 522 (2010) 536–554.
[14] R. Todeschini, A. Baccini, Handbook of bibliometric indicators: quantitative tools for studying and evaluating research, Wiley-VCH, Weinheim, Germany, 2016.
[15] O. Kinouchi, L. D. Soares, G. C. Cardoso, A simple centrality index for scientific social recognition, Physica A: Statistical Mechanics and its Applications 491 (2018) 632–640.
[16] O. Kinouchi, G. C. Cardoso, The k-index and the hubs of science., European heart journal 39 (2018) 3489.