P-INDEX - a fair alternative to H-INDEX

Adrian Horzyk [1]*

[1] Department of Automatics and Biomedical Engineering, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, Poland, 30-059 Cracow, Mickiewicza Av. 30.

*Correspondence to: horzyk@agh.edu.pl, phone: +48506084462

Abstract: The popular Hirsch index is used as a criterion for evaluating the achievements of scientists, departments, universities, and countries. Unfortunately, this index is often unfair in view of various scientific disciplines, can easily be manipulated, so its usefulness is undermined by unfair competition. To address these problems, this paper introduces a set of new popularity indices which evaluate the realistic popularity of a publication or a scientist better than the Hirsch index, by incorporating the number of different and non-repeating citing authors. These indices can rise only when citations from new authors appear. The popularity index is based on the scientist's most cited papers represented by the biggest number of non-repeating citing authors. Self- and duplicated citations of the same publication by the same authors are not taken into account. The p-index can also be applied to measure the productivity, popularity, and impact of a group of scientists as well as a scholarly journal. They can be treated as post-reviews of publications, because the cited papers or books would need to be positively reviewed by the citing authors in order to influence the popularity indices. Popularity indices are hard to manipulate or artificially increase, because the citing authors are considered only once. Thus, the p-index can be used as a fair alternative to the h-index.

Keywords: Hirsch index, h-index, impact factor, IF, popularity index, p-index.

1. Definition and purpose of h-index

The h-index is based on the distribution of citations that a researcher received for his or her publications. It reflects both the number of publications and the number of citations per publication. Hirsch wrote: "*A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p − h) papers have no more than h citations each* [14, 27]." It means that a scholar with an index of *h* has published *h* papers each of which has been cited in other papers at least h times. This index is designed to improve upon simpler measures such as the total number of citations or publications. Hirsch wrote: "*I propose the index h … as a useful index to characterize the scientific output of a researcher.*" [14] The h-index attempts to measure both the productivity and impact of the published work of a scientist or scholar. The index is calculated using a set of a scientist's most cited papers and the number of citations that they have received in other publications [27].

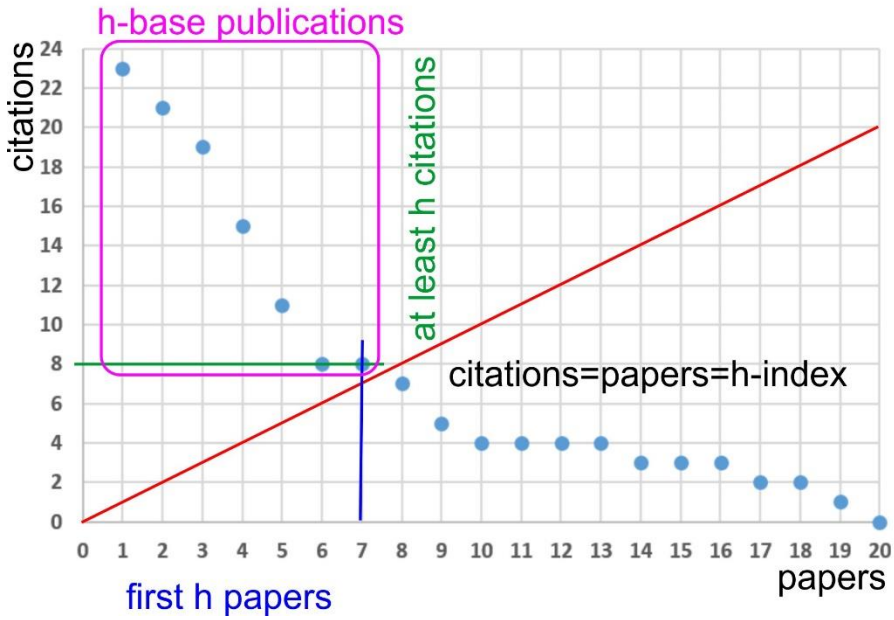

Fig. 1. A graphical example of computing the Hirsch index (h-index).

The h-index provides an alternative to more traditional journal impact factor metrics in the evaluation of the scientific impact of the work of particular researchers or research centers. The h-index grows as citations accumulate; the most often cited articles contribute to the h-index. Hirsch has shown that the h-index can be used to predict whether a scientist has won honors such as the Nobel Prize or National Academy membership [15]. The h-index can be manually determined using citation databases or using automatic tools and available calculators (e.g. in Scopus and Web of Knowledge databases). Harzing's Publish or Perish program provides h-indices calculated using Google Scholar entries.

Hirsch intended the h-index to address the main disadvantages of other bibliometric indicators, such as total number of papers or total number of citations [14, 15]. He claimed that the total number of papers does not account for the quality of scientific publications, while the total number of citations can be disproportionately affected by participation in a single publication of major influence. Low-quality methodological papers proposing new techniques or methods can generate a larger number of citations than high-quality publications. The h-index is intended to measure simultaneously the quality and quantity of scientific output [14].

2. Drawbacks of h-index

There is much misleading information about a scientist's output and the drawbacks of using the h-index to evaluate and compare the achievements of individual scientists, research groups, institutions, or countries [8, 26, 27]:

- The h-index works properly only for comparing scientists working in the same field [14], because the *h*-index does not take into account the typical number of citations across other fields – different fields or journals traditionally use different numbers of citations [1, 16, 20, 22].
- The *h*-index does not differentiate between books and articles, giving them the same mark, which makes it difficult to compare scholars in fields that produce more book output such as the humanities.
- The currently used *h*-index disregards the number of authors of a citing paper, so the popularity of the paper is not precisely emphasized. In the original paper, Hirsch suggested partitioning citations also among co-authors [14, 15].
- The *h*-index discards the information about author affiliation, which sometimes is significant [23, 28].

- The *h*-index is bounded by the time and the total number of publications, so discoveries of scientists with a short career are belittled regardless of their importance.
- The *h*-index does not consider the context of citations and equally treats criticizing and appreciative citations of a paper.
- The *h*-index can be manipulated through self-citations [3, 8, 11].

Scientists and scholars have proposed various ways to modify the *h*-index in order to emphasize other important features of publications [2, 5, 6, 17, 18, 24]. Although many variants have proliferated, most of the proposals do not differ significantly from the original *h*-index and they remain highly correlated with it. Here is a number of examples:

- An individual h-index normalized by the average number of co-authors in the h-score, where the number of citations is divided by the number of authors before ordering the papers and obtaining the h-index, as originally suggested by Hirsch [6]. Unfortunately, the distribution of the h-index further depends on the field of study.
- The **m-index** (m-quotient) is defined as the h-index divided by the number of years elapsed since the first published paper of a scientist in order to compare scientists regardless of time of publishing [4, 13, 14].
- The **g-index** is the h-index for an average count of citations [10].
- The **s-index** accounts for the non-entropic distribution of citations [25].
- The **c-index** considers the quality of citations in terms of the collaboration distance between the citing and cited authors. A scientist has *c*-index equal *n* if *n* of his or her *N* citations are from authors which are at collaboration distance at least *n*, and the other ($N - n$) citations are from authors which are at collaboration distance at most *n*. The collaboration distance is the distance between two nodes (representing scientists) in a collaboration graph, in which two distinct participants are joined by an edge whenever there is a collaborative relationship between them [7].
- The ***e*-index** takes into account the square root of surplus citations for the *h*-set beyond h^2, which is especially useful for highly cited scientists and for comparing those with the same *h*-index [9, 29].

3. Definition and purpose of popularity indices

This paper introduces popularity indices for evaluating scientific publications and their authors on the basis of the number of non-repetitive citing authors in view of a given publication list (an example is given in Tab. 1). In general, popularity is defined as liked, enjoyed, accepted, followed, used, or done by many people, frequently encountered or widely accepted, cited, related to, or coming from most of the people in a country, society, or group [21]. The popularity mentioned in this paper is defined by the number of other authors of scientific papers that cite the papers of a given author. The repetitions of citing the same paper by the same author are not taken into account because repetitions do not raise the real popularity of this paper. Frequent citation of the same paper by the same authors raises its impact factor (IF), and influence Hirsch index (h-index), but not the popularity index (p-index). Only new different authors can significantly increase the popularity indices of an author, publication, journal, department, university, or country.

3.1. Publication popularity index (pp-index)

The publication popularity index (pp-index) is defined as the total number of non-repeating authors that have cited a given publication at least once. All authors are counted exactly once (Tab. 1) and self-citations are omitted. It means that this publication is known to all of the citing authors. Popular publications of high scientific value are usually cited by various authors. The pp-indices indicate the real popularity of the publications. Thus, the pp-indices express the real worth of publications via citations of other citing authors.

Tab. 1. Computing of pp-indices for the example publications and their authors. The citing authors (in bold) are counted up only once for each publication. Self-citing authors are excluded from this count.

Title of the cited paper	Authors of the cited paper	No of all citing papers	No of citations without self-citation	No of non-repeating citing authors (pp-index)	Citing Authors	Citing papers
Self-optimizing neural networks	Horzyk, A; Tadeusiewicz, R	21	10	21	**Cpalka, Krzysztof; Zalasinski, Marcin; Rutkowski, Leszek**	New method for the on-line signature verification based on horizontal partitioning
					Cpalka, K.; **Lapa, K.; Przybyl, A.**; et al.	A new method for designing neuro-fuzzy systems for nonlinear modelling with interpretability aspects
					Cpalka, Krzysztof; Zalasinski, Marcin	On-line signature verification using vertical signature partitioning
					Lapa, Krystian; Zalasinski, Marcin; Cpalka, Krzysztof	A New Method for Designing and Complexity Reduction of Neuro-fuzzy Systems for Nonlinear Modelling
					Nowak, Bartosz A.; Nowicki, Robert K.; Mleczko, Wojciech K.	A New Method of Improving Classification Accuracy of Decision Tree in Case of Incomplete Samples
					Lorent, Anna; Knas, Michal; Dobosz, Piotr	Neuronal Model-Based Image Reconstruction from Projections Method
					Lapa, Krystian; Przybyl, Andrzej; Cpalka, Krzysztof	A New Approach to Designing Interpretable Models of Dynamic Systems
					Dudek-Dyduch, Ewa; Tadeusiewicz, Ryszard; Horzyk, Adrian	Neural network adaptation process effectiveness dependent of constant training data availability
					Dudek-Dyduch, Ewa; Horzyk, Adrian	The influence of training data availability time on effectiveness of ANN adaptation process
					Horzyk, Adrian	Introduction to constructive and optimization aspects of SONN-3
					Horzyk, Adrian	Self Optimizing Neural Networks SONN-3 for Classification Tasks
					Swiercz, W; Cios, KJ; Staley, K; et al.	A new synaptic plasticity rule for networks of spiking neurons
					Tadeusiewicz, Ryszard; **Izworski, Andrzej**	Learning in neural network - Unusual effects of "artificial dreams"
					Horzyk, A	Unsupervised clustering using self-optimizing neural networks
					Horzyk, A; Dudek-Dyduch, E	Effectiveness of artificial neural networks adaptation according to time period of training data acquisition
					Horzyk, A	Interval Basis Neural Networks
					Gorny, A; Tkacz, M	Using artificial neural networks for processing data gained via OpenDAP and consolidated from different databases on distributed servers
					Kotlarz, P; Kotulski, Z	On application of neural networks for S-boxes design
					Horzyk, A	A new extension of self-optimizing neural networks for topology optimization
					Horzyk, A; Tadeusiewicz, R	Comparison of plasticity of self-optimizing neural networks and natural neural networks
					Tadeusiewicz, R; Izworski, A; **Bulka, J**; et al.	Unusual effects of "artificial dreams" encountered during learning in neural network
Neural network adaptation process effectiveness dependent of constant training data availability	Dudek-Dyduch, E; Tadeusiewicz, R; Horzyk, A	5	4	7	**Glowacz, Adam; Glowacz, Andrzej; Glowacz, Zygfryd**	Recognition of monochrome thermal images of synchronous motor with the application of quadtree decomposition and backpropagation neural network
					Glowacz, W.	Diagnostics of induction motor based on spectral analysis of stator current with application of backpropagation neural network
					Szaleniec, Maciej	Prediction of enzyme activity with neural network models based on electronic and geometrical features of substrates
					Horzyk, Adrian	Information Freedom and Associative Artificial Intelligence
					Dabrowski, Artur; Jach, Anna; Kapitaniak, Tomasz	Application of artificial neural networks in parametrical investigations of the energy flow and synchronization

3.2. Author popularity index (ap-index)

The author popularity index (ap-index) is defined as the total number of other authors that have cited any publication written by a given author at least once. As in the case of the pp-index, each citing author is counted only once. This index indicates how many other authors are familiar with publications, appreciate the scientific work, and use research results of a given author. Valuable research results described in publications are usually cited by many authors, which this index expresses for a cited author. The more popular a cited author in a field and scientific community, the higher the ap-index. The citations of citing authors also appreciate a given author and reproduce a quality measure in a similar fashion as a reviewing process. Each cited publication can be treated as positively post-reviewed by the citing authors.

3.3. Popularity index (p-index)

The popularity index (p-index) is defined as the number of a given author's publications, which have a pp-index at least equal to this number (Fig. 2). This index is comparable to the h-index and is similarly calculated but it is determined on the basis of all citing authors, not citations of publications. It can be raised only when many authors are familiar with and cite many publications of a given author. This is a natural process correlated with popularity of work and publications of a given cited author. The main difference is that other citing authors are taken into account only once for a given publication. The same citing author can be counted many times across different publications

but only once for a single publication of a given author. This index expresses the popularity of the author's papers in a community of other researchers. Thus, the p-index can be used for evaluation of scientific work, author's publications, value of scientific achievements, productivity, and author's popularity. The growth of p-index is effectively stopped in case of frequent citations of the same publications by the same authors. This eliminates one of the main drawbacks of the h-index.

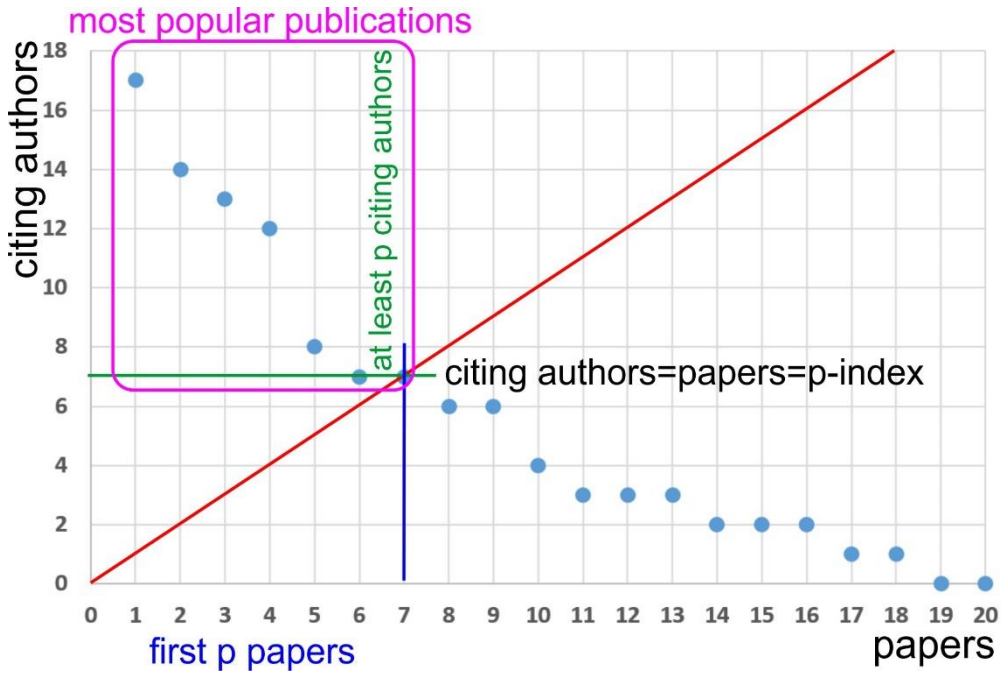

Fig. 2. An example of determining the popularity index (p-index) of an author.

Tab. 2. An example of citations and non-repeating citing authors for 20 publications of an analyzed author.

Papers	Citations	Non-repeating citing authors
P1	8	12
P2	15	14
P3	21	13
P4	19	7
P5	3	6
P6	23	17
P7	2	1
P8	11	8
P9	4	3
P10	3	2
P11	4	6
P12	7	4
P13	2	3
P14	1	0
P15	4	2
P16	5	3
P17	4	1
P18	0	0
P19	3	2
P20	8	7

Papers	Citations
P6	23
P3	21
P4	19
P2	15
P8	11
P1	8
P20	8
P12	7
P16	5
P9	4
P11	4
P15	4
P17	4
P5	3
P10	3
P19	3
P7	2
P13	2
P14	1
P18	0
h-index	**7**

Papers	Non-repeating citing authors
P6	17
P2	14
P3	13
P1	12
P8	8
P4	7
P20	7
P5	6
P11	6
P12	4
P9	3
P13	3
P16	3
P10	2
P15	2
P19	2
P7	1
P17	1
P14	0
P18	0
p-index	**7**

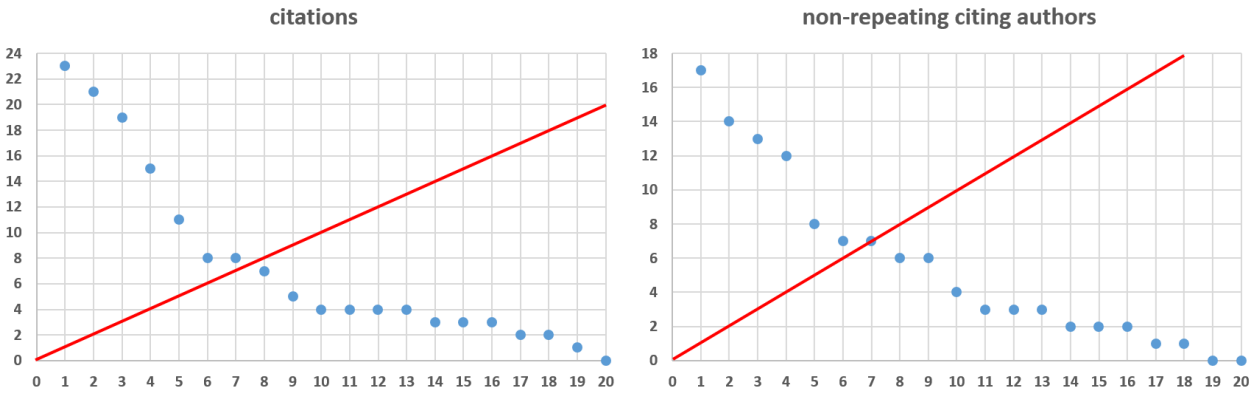

Fig. 3. A comparison of citations and non-repeating citing authors for a given set of 20 publications listed in Tab. 2 used to compute the h-index and p-index.

Table 2 shows a sample set of 20 publications of a cited author, their numbers of citations (without self-citations), and numbers of other different citing authors (without multiple citations of the same publications by the same authors). The frequently cited publications can sometimes have few citing authors. In comparison to h-index, which measures the worth of an author's publications using citation numbers, the p-index measures their worth using numbers of different citing authors that have appreciated the worth of the publications via citing them. Popularity index can be treated as a post-review of the cited publications in view of the citing authors – the real value of such publications.

Figure 3 compares the results obtained for an author for a given set of 20 publications listed in Tab. 2. The citation graph is used to calculate the h-index; the citing authors' graph is used to calculate the p-index. In this case the graphs are very similar. When citations are not artificially raised, the h-index and p-index are very similar. Otherwise, the p-indices will be more accurate in view of real popularity and value of the cited publications.

3.4. Journal popularity index (jp-index) and factor (jp-factor)

The journal popularity index (jp-index) is the number of the authors that have at least once cited any of the papers of a journal. Each author is counted up only once regardless of how many papers of this journal he has cited. This index expresses the popularity of the journal by the citing authors. Only new different authors can raise the popularity of a journal.

The journal popularity factor (jp-factor or JPF) is defined as the ratio of the sum of the pp-indices of all papers published in the given journal to the total number of papers.

3.5. Other popularity indices and factors

The popularity indices and factors can be similarly used to measure the popularity of groups, centres, departments, universities, and countries. They are universal and can be treated as post-reviews of the research results introduced in publications because usually only valuable results are cited and compared. Few publications cite worthless results and if they do that, they are usually not further cited. Thus, popularity indices can be successfully used as an objective and impartial appraisal of a given author (scientific group, department, university, or country) or as a post-review of his/her research and its results made by all citing authors of their cited publications.

4. Conclusion

This paper introduced new popularity indices that can be used to indicate the real popularity of authors and their publications as well as journals and scientific centers, departments, universities, or countries. These indices use the

number of non-repeating citing authors instead of the number of citations in order to express real popularity and value of cited publications. These indices can be used as a kind of post-review of publications because the cited papers or books would have to be positively reviewed first by the citing authors that have decided to use them in their scientific study. The true value of the popularity indices lies in their considering each citing author only once for each publication or each cited author. Thus, these indices express real popularity and worth in view of all citing authors. This feature of popularity indices eliminates one of the main drawbacks of the h-index that can be artificially raised by many citations made by the same citing authors. The popularity indices would be hard to manipulate because they rely on the community of citing authors; they naturally grow with the value of published scientific work and its results. Thus, popularity indices – in particular the p-index – can be used as a fair alternative to the h-index. The p-index does not remove all drawbacks of the h-index, but enables to better appreciate work of scientists accordingly to the number of non-repeating citing authors that have appreciated results of their work.

References

[1] Alonso, S. (2009). H-index: a review focused on its variants, computation and standardization for different scientific fields. *J. Informetrics* **3**, 273–289.

[2] Anderson, T. R., Hankin, R. K. S., & Killworth, P. D. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics* **76** (3): 577–588, doi:10.1007/s11192-007-2071-2.

[3] Bartneck, Christoph, & Kokkelmans, Servaas (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics* **87** (1): 85–98, doi:10.1007/s11192-010-0306-5.

[4] Von Bohlen, & Halbach, O. (2011). „How to judge a book by its cover? How useful are bibliometric indices for the evaluation of 'scientific quality' or 'scientific productivity'? Annals of Anatomy **193** (3): 191–6, doi:10.1016/j.aanat.2011.03.011.

[5] Baldock, C., Ma, R. M. S., & Orton, C. G. (2009). „The h index is the best measure of a scientist's research productivity. *Medical Physics* **36** (4): 1043–1045, doi:10.1118/1.3089421.

[6] Batista, P. D. (2006). Is it possible to compare researcher with different scientific interests? *Scientometrics* **68** (1), 179–189, doi:10.1007/s11192-006-0090-4.

[7] Bras-Amorós, M., Domingo-Ferrer, J., & Torra, V. (2011). „A bibliometric index based on the collaboration distance between cited and citing authors. *Journal of Informetrics* **5** (2): 248–264, doi:10.1016/j.joi.2010.11.001.

[8] Czarnecki, L., Kazmierkowski, M.P., & Rogalski, A. (2013). Doing Hirsch proud; shaping H-index in engineering sciences, Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. **61**, No. 1, DOI: 10.2478/bpasts-2013-0001.

[9] Dodson, M.V. (2009). „Citation analysis: Maintenance of h-index and use of e-index. *Biochemical and Biophysical Research Communications* **387** (4): 625–6, doi:10.1016/j.bbrc.2009.07.091.

[10] Egghe, Leo (2013). „Theory and practise of the g-index. *Scientometrics* **69**: 131, doi:10.1007/s11192-006-0144-7.

[11] Ferrara, Emilio, & Romero, Alfonso (2013). „Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index. *Journal of the American Society for Information Science and Technology* **64** (11): 2332–2339, doi:10.1002/asi.22976.

[12] Garfield, E. (1955). Citation indexes to science: a new dimension in documentation through association of ideas, *Science* **122** (3159), 108–111.

[13] Harzing, A.W., & van der Wall, R. (2007). A google scholar h-index for journals: a better metric to measure journal impact in economics and business, www.harzing.com.

[14] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output, *PNAS* **102** (46): 16569-16572, doi:10.1073/pnas.0507655102.

[15] Hirsch, J. E. (2007). Does the h-index have predictive power? *PNAS* **104** (49): 19193–19198.

[16] Iglesias, J. E., & Pecharroman, C. (2007). Scaling the h-indexes for Different Scientific ISI Fields, *Scientometrics* **73** (1), 303–320.

[17] Jayant, S. Vaidya (2005). „V-index: A fairer index to quantify an individual's research output capacity. BMJ **331**: 339–c–1340–c.

[18] Katsaros, D., Sidiropoulos, A., & Manolopous, Y. (2007). Age Decaying H-Index for Social Network of Citations, in Proceedings of Workshop on Social Aspects of the Web Poznan, Poland, April 27.

[19] Kierzek, R. (2008). Polish science in Hirsch index, *Science Matters* **6–7**, 137.

[20] Lillquist, E., & Green, S. (2010). The discipline dependence of citation statistics, *Scientometrics* **84**, 749–768.

[21] *Merriam-Webster Dictionary*, http://www.merriam-webster.com/.

[22] Podlubny I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science, *Scientometrics* **64** (1), 95–99.
[23] Sekercioglu, H. Cagan (2008). Quantifying coauthor contributions. *Science* **322** (5900): 371, doi:10.1126/science.322.5900.371a.
[24] Sidiropoulos, Antonis, Katsaros, Dimitrios, & Manolopoulos, Yannis (2007). „Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics* **72** (2): 253–280, doi:10.1007/s11192-007-1722-z.
[25] Silagadze, Z. K. (2010). Citation entropy and research impact estimation. *Acta Phys. Polon. B* **41**: 2325–2333.
[26] Wendl, Michael (2007). „H-index: however ranked, citations need context. *Nature* **449** (7161): 403, doi:10.1038/449403b.
[27] Wikipedia: http://en.wikipedia.org/wiki/H-index#cite_note-11.
[28] Chun-Ting, Zhang (2009). A proposal for calculating weighted citations based on author rank. EMBO reports **10** (5): 416–7, doi:10.1038/embor.2009.74.
[29] Chun-Ting, Zhang (2009). „The e-Index, Complementing the h-Index for Excess Citations. In Joly, Etienne, PLoS ONE **4** (5): e5429, doi:10.1371/journal.pone.0005429.