h-index

From Wikipedia, the free encyclopedia

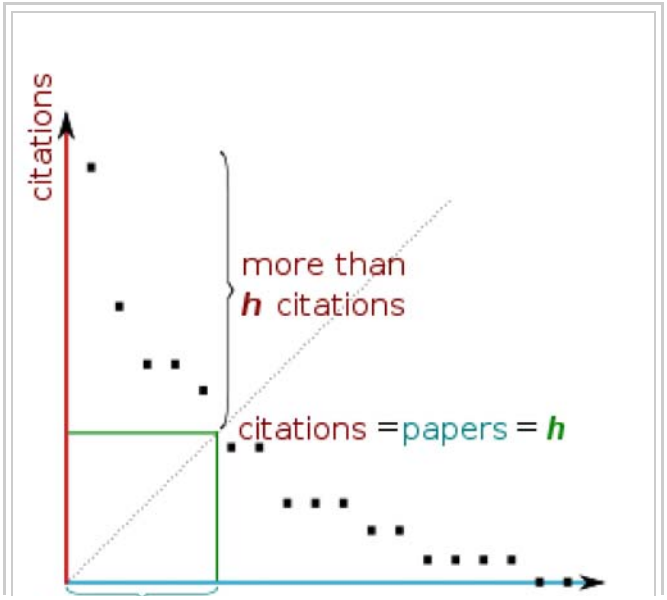
The ***h*-index** is an index that attempts to measure both the scientific productivity and the apparent scientific impact of a scientist. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other people's publications. The index can also be applied to the productivity and impact of a group of scientists, such as a department or university or country. The index was suggested by Jorge E. Hirsch, a physicist at UCSD, as a tool for determining theoretical physicists' relative quality[1] and is sometimes called the *Hirsch index* or *Hirsch number*.

Contents

Definition and purpose

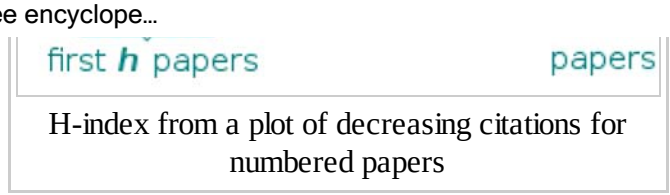
The index is based on the distribution of citations received by a given researcher's publications. Hirsch writes:

> *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 at most h citations each.*



In other words, a scholar with an index of *h* has published *h* papers each of which has been cited by others at least *h* times.[2] Thus, the *h*-index reflects both the number of publications and the number of citations per publication. The index is designed to improve upon simpler measures such as the total number of citations or publications. The index works properly only for comparing scientists working in the same field; citation conventions differ widely among different

fields.



H-index from a plot of decreasing citations for numbered papers

The *h*-index serves as an alternative to more traditional journal impact factor metrics in the evaluation of the impact of the work of a particular researcher. Because only the most highly cited articles contribute to the *h*-index, its determination is a relatively simpler process. Hirsch has demonstrated that *h* has high predictive value for whether a scientist has won honors like National Academy membership or the Nobel Prize. In physics, a moderately productive scientist should have an *h* equal to the number of years of service while biomedical scientists tend to have higher values. The *h*-index grows as citations accumulate and thus it depends on the 'academic age' of a researcher.

Hirsch suggested that, for physicists, a value for *h* of about 10–12 might be a useful guideline for tenure decisions at major research universities. A value of about 18 could mean a full professorship, 15–20 could mean a fellowship in the American Physical Society, and 45 or higher could mean membership in the United States National Academy of Sciences.[3] Little systematic investigation has been made on how academic recognition correlates with h-index over different institutions, nations and fields of study.

The h-index has been characterized in terms of three natural axioms by Woeginger .[4] The simplest of these three axioms states that by increasing the number of citations to a single article, the research index should not increase by more than 1.

Calculating *h*

The *h*-index can be manually determined using free Internet databases, such as Google Scholar or freeware such as Publish or Perish. Subscription-based databases such as Scopus and the Web of Knowledge provide automated calculators. Each database is likely to produce a different *h* for the same scholar, because of different coverage: Google Scholar has more citations than Scopus and Web of Science but the smaller citation collections tend to be more accurate. In addition, specific databases, such as Stanford Physics Information Retrieval System (SPIRES) can automatically calculate h-index for researchers working in High Energy Physics.

The topic has been studied in detail by Lokman I. Meho and Kiduk Yang.[5][6] Web of Knowledge was found to have strong coverage of journal publications, but poor coverage of high impact conferences. Scopus has better coverage of conferences, but poor coverage of publications prior to 1996; Google Scholar has the best coverage of conferences and most journals (though not all), but like Scopus has limited coverage of pre-1990 publications.[6] The exclusion of conference preprints is a problem for scholars in computer science, where conference preprints are considered an important part of the literature, but reflects common practice in most scientific fields where conference preprints are unrefereed and are accorded less weight in evaluating academic productivity. Google Scholar has also been criticized for producing "phantom citations," including gray literature in its citation counts, and failing to follow the rules of Boolean logic when combining search terms.[7] For example, the Meho and Yang study found that Google Scholar identified 53% more citations than Web of Knowledge and Scopus combined, but noted that most of the additional citations reported by Google Scholar were from low-impact journals or conference proceedings. It has been suggested that in order to deal with the sometimes wide variation in *h* for a single academic measured across the possible citation databases, that one could assume false negatives in the databases are more problematic than false positives and take the maximum *h* measured for an academic.[8]

Advantages

The *h*-index was intended to address the main disadvantages of other bibliometric indicators, such as total number of papers or total number of citations. Total number of papers does not account for the quality of scientific publications, while total number of citations can be disproportionately affected by participation in a single publication of major influence. The *h*-index is intended to measure simultaneously the quality and sustainability of scientific output, as well as, to some extent, the diversity of scientific research. The *h*-index is much less affected by methodological papers proposing successful new techniques, methods or approximations, which can be extremely highly cited.

Criticism

There are a number of situations in which *h* may provide misleading information about a scientist's output:[9]

- The *h*-index does not account for the number of authors of a paper. In the original paper, Hirsch suggested partitioning citations among co-authors. Even in the absence of explicit gaming, the *h*-index and similar indexes tend to favor fields with larger groups, e.g. experimental over theoretical.

- The *h*-index does not account for the typical number of citations in different fields. Different fields, or journals, traditionally use different numbers of citations.

- The *h*-index is bounded by the total number of publications. This means that scientists with a short career are at an inherent disadvantage, regardless of the importance of their discoveries. For example, Évariste Galois' *h*-index is 2, and will remain so forever. Had Albert Einstein died in early 1906, his *h*-index would be stuck at 4 or 5. This is also a problem for any measure that relies on the number of publications. As Hirsch indicated in the original paper, the index intended as a tool to evaluate researchers in the same stage of their careers. It is not meant as a tool for historical comparisons.

- The *h*-index does not consider the context of citations. For example, citations in a paper are often made simply to flesh out an introduction, otherwise having no other significance to the work. *h* also does not resolve other contextual instances: citations made in a negative context and citations made to fraudulent or retracted work. This is also a problem for regular citation counts.

- The *h*-index does not account for confounding factors such as "gratuitous authorship", the so-called Matthew effect, and the favorable citation bias associated with review articles. Again, this is a problem for all other metrics using publications or citations.

- The *h*-index has been found to have slightly less predictive accuracy and precision than the simpler measure of mean citations per paper.[10] However, this finding was contradicted by another study.[11]

- The *h*-index is a natural number and thus lacks discriminatory power. Ruane and Tol therefore propose a rational *h*-index that interpolates between *h* and *h*+1.[12]

Alternatives and modifications of the h-index

- Various proposals to modify the *h*-index in order to emphasize different features have been made.[13][14][15][16][17][18]

- An individual *h*-index normalized by the average number of co-authors in the *h*-core has been introduced by

Batista *et al.*[13] They also found that the distribution of the *h*-index, although it depends on the field, can be normalized by a simple rescaling factor. For example, assuming as standard the *h*s for biology, the distribution of *h* for mathematics collapse with it if this *h* is multiplied by three, that is, a mathematician with $h = 3$ is equivalent to a biologist with $h = 9$. This method has not been readily adopted, perhaps of its complexity. It might be simpler to divide citation counts by the number of authors before ordering the papers and obtaining the h-index, as originally suggested by Hirsch.

- The *m*-index is defined as h/n, where n is the number of years since the first published paper of the scientist [1]

See also

- Bibliometrics
- impact factor
- g-index
- h-b index
- Eddington number (cycling) An earlier metric of the same form.

References

1. ^ ***a*** ***b*** Hirsch, J. E. (15 November 2005). "An index to quantify an individual's scientific research output" (http://www.pnas.org/content/102/46/16569.full) . *PNAS* **102** (46): 16569–16572. doi:10.1073/pnas.0507655102 (http://dx.doi.org/10.1073%2Fpnas.0507655102) . PMID 16275915 (http://www.ncbi.nlm.nih.gov/pubmed/16275915) . PMC 1283832 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1283832) . http://www.pnas.org/content/102/46/16569.full. Retrieved 13 May 2010 (see also pdf format (http://www.pnas.org/content/102/46/16569.full.pdf+html)
2. ^ McDonald, Kim (8 November 2005). "Physicist Proposes New Way to Rank Scientific Output" (http://www.physorg.com/news7971.html) . PhysOrg. http://www.physorg.com/news7971.html. Retrieved 13 May 2010.
3. ^ Peterson, Ivars (December 2, 2005). "Rating Researchers" (http://www.sciencenews.org/view/generic/id/6824/title/Math_Trek__Rating_Researchers) . *Science News*. http://www.sciencenews.org/view/generic/id/6824/title/Math_Trek__Rating_Researchers. Retrieved 13 May 2010.
4. ^ Woeginger, GJ (2008). "An axiomatic characterization of the Hirsch-index". *Mathematical Social Sciences* **56**: 224–232. doi:10.1016/j.mathsocsci.2008.03.001 (http://dx.doi.org/10.1016%2Fj.mathsocsci.2008.03.001) .
5. ^ Meho, L. I.; Yang, K. (2007). "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar". *Journal of the American Society for Information Science and Technology* **58** (13): 2105–2125. doi:10.1002/asi.20677 (http://dx.doi.org/10.1002%2Fasi.20677) .
6. ^ ***a*** ***b*** Meho, L. I. and Yang, K (23 December 2006). "A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar" (http://arxiv.org/ftp/cs/papers/0612/0612132.pdf) . *arXiv.org, Computer Science, cs/0612132, 23*. arXiv.org. http://arxiv.org/ftp/cs/papers/0612/0612132.pdf. Retrieved 13 May 2010 (preprint of paper published as 'Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar', in *Journal of the American Society for Information Science and Technology*, Vol. **58**, No. 13, 2007, 2105-2125)
7. ^ Jacsó, Péter (2006). "Dubious hit counts and cuckoo's eggs". *Online Information Review* **30** (2): 188–193. doi:10.1108/14684520610659201 (http://dx.doi.org/10.1108%2F14684520610659201) .
8. ^ Sanderson, Mark (2008). "Revisiting *h* measured on UK LIS and IR academics". *Journal of the American Society for Information Science and Technology* **59** (7): 1184–1190. doi:10.1016/j.mathsocsci.2008.03.001 (http://dx.doi.org/10.1016%2Fj.mathsocsci.2008.03.001) .

9. ^ Wendl, Michael (2007). "H-index: however ranked, citations need context". *Nature* **449** (7161): 403. doi:10.1038/449403b (http://dx.doi.org/10.1038%2F449403b) . PMID 17898746 (http://www.ncbi.nlm.nih.gov/pubmed/17898746) .
10. ^ Sune Lehmann, Andrew D. Jackson, Benny E. Lautrup (2006). "Measures for measures". *Nature* **444** (7122): 1003–4. doi:10.1038/4441003a (http://dx.doi.org/10.1038%2F4441003a) . PMID 17183295 (http://www.ncbi.nlm.nih.gov/pubmed/17183295) .
11. ^ Hirsch J. E. (2007). "Does the h-index have predictive power?" (http://www.pnas.org/content/104/49/19193.full) . *PNAS* **104** (49): 19193–19198. doi:10.1073/pnas.0707962104 (http://dx.doi.org/10.1073%2Fpnas.0707962104) . PMID 18040045 (http://www.ncbi.nlm.nih.gov/pubmed/18040045) . PMC 2148266 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=2148266) . http://www.pnas.org/content/104/49/19193.full. (pdf version (http://www.pnas.org/content/104/49/19193.full.pdf+html))
12. ^ Frances Ruane & Richard S. J. Tol (2008). "Rational (successive) h -indices: An application to economics in the Republic of Ireland". *Scientometrics* **75** (2): 395–405. doi:10.1007/s11192-007-1869-7 (http://dx.doi.org/10.1007%2Fs11192-007-1869-7) .
13. ^ ***a b*** Batista P. D. et al. (2006). "Is it possible to compare researchers with different scientific interests?". *Scientometrics* **68** (1): 179–189. doi:10.1007/s11192-006-0090-4 (http://dx.doi.org/10.1007%2Fs11192-006-0090-4) .
14. ^ 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 (http://dx.doi.org/10.1007%2Fs11192-007-1722-z) .
15. ^ Jayant S Vaidya (December 2005). "V-index: A fairer index to quantify an individual's research output capacity" (http://bmj.com/cgi/eletters/331/7528/1339-c#123188) . *BMJ* **331**: 339-c-1340-c. http://bmj.com/cgi/eletters/331/7528/1339-c#123188.
16. ^ Katsaros D., Sidiropoulos A., Manolopous Y., (2007), Age Decaying H-Index for Social Network of Citations (http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-245/paper3.pdf) in Proceedings of Workshop on Social Aspects of the Web Poznan, Poland, April 27, 2007 (http://ceur-ws.org/Vol-245/)
17. ^ Anderson, T.R.; Hankin, R.K.S and 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 (http://dx.doi.org/10.1007%2Fs11192-007-2071-2) .
18. ^ Baldock, C.; Ma, R.M.S and 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 (http://dx.doi.org/10.1118%2F1.3089421) . PMID 19472608 (http://www.ncbi.nlm.nih.gov/pubmed/19472608) .

Further reading

- Ball, Philip (2005). "Index aims for fair ranking of scientists". *Nature* **436** (7053): 900. doi:10.1038/436900a (http://dx.doi.org/10.1038%2F436900a) . PMID 16107806 (http://www.ncbi.nlm.nih.gov/pubmed/16107806) .
- Kelly, C. D.; Jennions, M. D. (2006). "The h index and career assessment by numbers". *Trends Ecol. Evol. (Amst.)* **21** (4): 167–70. doi:10.1016/j.tree.2006.01.005 (http://dx.doi.org/10.1016%2Fj.tree.2006.01.005) . PMID 16701079 (http://www.ncbi.nlm.nih.gov/pubmed/16701079) .
- Lehmann, S.; Jackson, A. D.; Lautrup, B. E. (2006). "Measures for measures". *Nature* **444** (7122): 1003–4. doi:10.1038/4441003a (http://dx.doi.org/10.1038%2F4441003a) . PMID 17183295 (http://www.ncbi.nlm.nih.gov/pubmed/17183295) .
- 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 (http://dx.doi.org/10.1007%2Fs11192-007-1722-z) .

- Soler, José M. (2007). "A rational indicator of scientific creativity". *Journal of Informetrics* **1** (2): 123–130. doi:10.1016/j.joi.2006.10.004 (http://dx.doi.org/10.1016%2Fj.joi.2006.10.004) .
- Symonds, M. R.; *et al.* (2006). "Gender differences in publication output: towards an unbiased metric of research performance" (http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1762413) . *PLoS ONE* **1**: e127. doi:10.1371/journal.pone.0000127 (http://dx.doi.org/10.1371%2Fjournal.pone.0000127) . PMID 17205131 (http://www.ncbi.nlm.nih.gov/pubmed/17205131) .
- Taber, Douglass F. (2005). "Quantifying Publication Impact". *Science* **309** (5744): 2166a. doi:10.1126/science.309.5744.2166a (http://dx.doi.org/10.1126%2Fscience.309.5744.2166a) . PMID 16195445 (http://www.ncbi.nlm.nih.gov/pubmed/16195445) .
- Woeginger, Gerhard j. (2008). "An axiomatic characterization of the Hirsch-index". *Mathematical Social Sciences* **56** (2): 224–232. doi:10.1016/j.mathsocsci.2008.03.001 (http://dx.doi.org/10.1016%2Fj.mathsocsci.2008.03.001) .

External links

Computing the h-index

- Online computation of H and G index using Google Scholar: 1. ResEval tool by University of Trento, Italy (http://reseval.org/) , 2. A tool by INRIA Lille, France (http://interaction.lille.inria.fr/~roussel/projects/scholarindex/index.cgi) , 3. QuadSearch tool (http://quadsearch.csd.auth.gr/index.php?lan=1&s=2) , 4. CIDS tool discerning self-citations (http://cids.fc.ul.pt/) ,
- Scripts: 1. MATLAB script (http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=9710&objectType=file) to compute the h-index, 2. Publish or Perish (http://www.harzing.com/pop.htm) calculates various statistics, including the h- and g-index. This service requires downloading a program, which is available in PC and Linux formats (no Mac format),
- Browser Extensions: 1. Scholarometer (http://scholarometer.indiana.edu/) , a cross-platform browser extension, leverages the wisdom of crowds to analyze and compare citation impact, h-index based on Google Scholar data, with user-provided discipline tags, 2. A lightweight Firefox extension (https://addons.mozilla.org/en-US/firefox/addon/45283) , which computes citation indices on the fly when querying Google Scholar.
- h-index and variants (http://sci2s.ugr.es/hindex/) a website that introduces a comprehensive review on the h-index and related indices.

Lists of h-indices

- A long list of chemists (http://www.rsc.org/chemistryworld/News/2007/April/23040701.asp) with high h-index values
- The H-index for economists (http://ideas.repec.org/top/top.person.hindex.html) .
- The H-index for computer science (http://www.cs.ucla.edu/~palsberg/h-number.html)
- H values for Stanford p-chem professors (http://blog.everydayscientist.com/?p=10) from "The Everyday Scientist" (http://blog.everydayscientist.com/)
- h-index for Journals and Countries (http://www.scimagojr.com/)
- A paper with, among other things, the H values of the top Alzheimer's researchers

(http://iospress.metapress.com/content/v932x18k23300844/fulltext.html)
- Citation indices of Estonian/Latvian/Finnish computer scientists (http://research.cyber.ee/~lipmaa/cites/)
- A web site with the H values of graduate programs in philosophy in the US (http://el-prod.baylor.edu/certain_doubts/?page_id=774)

Retrieved from "http://en.wikipedia.org/wiki/H-index"
Categories: Bibliometrics | Academic publishing | Index numbers

---