PHYS 395 Research Methods - Lecture 4

Search for scientific information, bibliographic databases, scientometics, citations, Hirsch-index

Sergiy Bubin

Department of Physics
Nazarbayev University

Keeping up with the scientific literature

- To build up expertise/knowledge in a (new) field one must resort to original works
- To push frontiers, researchers must be aware of what is going on (i.e. where the frontier is)
- To contextualize one's work and find new applications

Good old days

Search for scientific information

- The amount of publications (and other information) has been growing exponentially
- Keeping up with the literature is a daunting task
- Academics/researchers tend to be busy with other things - experiments, writing proposals, teaching, administrative duties, etc

Bibliographic databases

- Non-electronic bibliographic databases has been around for a while
- They typically included the titles and abstracts of papers
- Some were comprehensive (e.g. Science Abstracts, Реферативный журнал), others collected information in a specific field (e.g. Engineering Index, Chemical Abstracts)

Modern electronic bibliographic databases

There are many. However, there are three giants that have become particularly popular

- Web of Science by Clarivative Analytics
- Scopus by Elsevier
- Google Scholar by Google

They implement various features and services which go well beyond indexing newly published papers

Other popular resources and databases

- PubMed
- Inspire
- NASA-ADS
- Microsoft Academic Search
- ReseachGate

arXiv

- Electronic preprint repository (moderated, but not fully peer-reviewed)
- Started as a physics archive at LANL, but later was expanded to other fields and is now managed by Cornell
- Has reached the milestone of 1 million preprints in 2014
- Historically papers were submitted in TeX format

Citations as a measure of impact

The analysis of citation statistics allows to quantify the impact a particular research or researcher is making.

- Article-level citations
- Author-level citations
- Using citation metric as a measure of quality can be controversial, yet it is gaining popularity

Journal impact factor

The impact factor (IF) is a measure of the frequency with which the average article in a journal has been cited in a particular year. It is used to measure the importance or rank of a journal by calculating the times it's articles are cited

$$IF_y = \frac{Citations_{y-1} + Citations_{y-2}}{Publications_{y-1} + Publications_{y-2}} \tag{1}$$

Here y denotes a year (e.g. 2015). Impact factors change over time reflecting changes in popularity and visibility of journals.

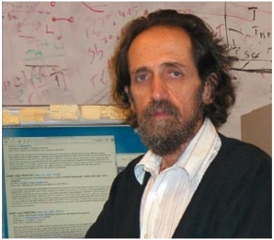
Impact factors of some journals

Journal title	IF (2018)
CA-A Cancer Journal for Clinicians	223.679
New England Journal of Medicine	70.670
Nature	43.070
Science	41.063
Reviews of Modern Physics	38.296
Nature Photonics	
31.583	
Physical Review Letters	9.227
Physical Review A	2.907
Physical Review B	3.736
Journal of the Korean Physical Society	0.630

h-index

The **Hirsch index**, or simply the **h-index**, is a metric that measures the publication productivity of a researcher and the citation impact of his/her works.

This index was introduced [Proc. Natl. Acad. Sci. USA 102, 16569 (2005)] by Jorge Eduardo Hirsch, a professor of condensed-matter physics at UCSD, in 2005



h-index

h-index is the number of papers (h) that have received at least h citations. As an example, an h index of 10 means that among all publications by one author, 10 of these publications have received at least 10 citations each.

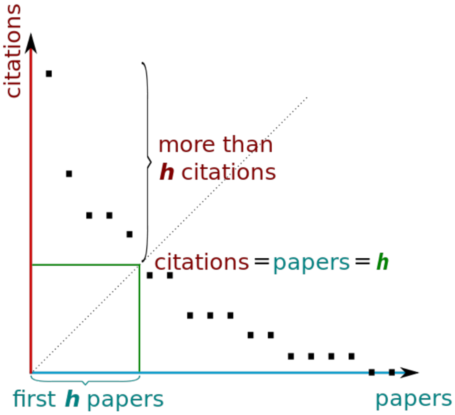


image: wikipedia.org

Good things about the h-index

- Simplicity – given by a single number
- Relies on citations to your papers, not the reputation of journals, which is a truer measure of quality
- Not skewed dramatically by a single well-cited, influential paper (unlike total number of citations would be)
- Not increased by a large number of poorly cited papers (unlike total number of papers would be)
- May also be used to compare not just individuals, but departments, programs or any other group of researchers

Hirsch argued that the h-index is preferable to other single-number criteria, such as the total number of papers, the total number of citations and citations per paper.

Weaknesses of the h-index

- A single number can never give an accurate representation of a complex thing
- It does not account for variations in average number of publications and citations in various fields (some traditionally publish and cite less than others)
- It does not fully reflect accomplishments of a scientist who authors relatively low number of few seminal papers with extraordinarily high citation counts
- It ignores the number and position of authors on a paper
- It has relatively low resolution – many scientists end up in the same range since it gets increasingly difficult to increase h the higher it gets

Weaknesses of the h-index

Lies, dams lies, statistics – the power of numbers, particularly the use of statistics, may be deceiving sometimes

Other scholar indicies

There certainly exist other indicies that measure productivity:

- m-quotient = h-index/n, where n is the number of years since the first published paper of the scientist
- g-index – the largest number such that the top g articles received, in total, at least g^2 citations. It rewards exceptionally well-cited articles
- e-index – the square root of the sum of the "exces" citations in the papers that contributed to the h-index.
- Erdős number (more like humor)