H5-index: rationale, design, features, advantages

Anurag Acharya

Circa 1990

Print was the dominant format

What your shelves could hold was what you could read

Libraries needed to maximize "ROI" within shelf-space & cost budgets

Every article had to be stored, took up shelf-space

Impact Factor provided estimated utility per unit shelf-space (~15 page article)

In effect, average number of citations per inch

Circa 1990 - II

Wide distribution for journals with high estimated utility for each shelf-row

    A few journals reached large audiences (most libraries)

    Others had a limited audience (few libraries)

Authors tried to publish in widely distributed journals

    Reach more readers, more likely to help build reputation

Tenure committees counted articles in "core" journals

Circa 2017

Almost all journals are online

    Shelf-space is no longer a constraint, citations/inch no longer needed

All journals have the same audience size

    Audience size no longer useful input for submission decisions

Per article citations stats are easy to find

    Tenure committees have direct evidence of articles' impact

Circa 2017 - II

Relevance ranking allows readers to focus on articles they want

    Readers search for & read articles, not journals

    Unlike browse/time-based ranking, presence of other articles has no impact

For libraries: success is subscriptions being used

    Journals with more "well-known" papers more likely to be used

Circa 2017 - III

For authors: success is reputation in community

What are examples of success in this journal?

How many papers in this journal have been successful?

What is the community of authors I will be joining?

Publication decisions are aspirational, you shoot for success

No one shoots for writing the "average" paper

What should today's journal metric look like?
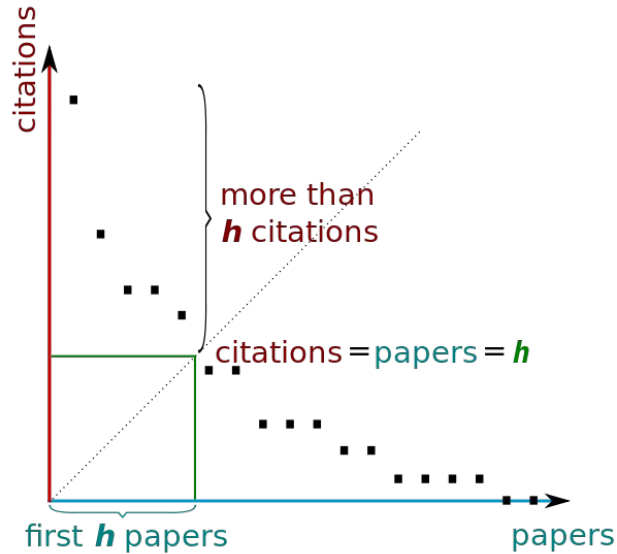
Focus on the "successful" papers

      For authors: how big is the successful community in this journal?

      For libraries: how many papers in this journal are likely to be widely read?

H-index

Focuses on successful papers - upper end of citation distribution

Journal with h-index of 50 has at least 50 papers with 50 citations each

H5-index

H-index for all papers over a 5 year period

Picked 5 years to cover citation rates of as many fields as possible

Wide variance in citation rates across fields

HSS slower than STM, experimental slower than theoretical

Some fields centered around bi-annual conferences

H5-median

H-index ignores distribution of citations in h-core

H-median tries to bring back a measure of this distribution

    Median of citation counts for all articles in journal's h-core

Used to break ties between journals with same h5-index

Scholar Metrics: inclusion

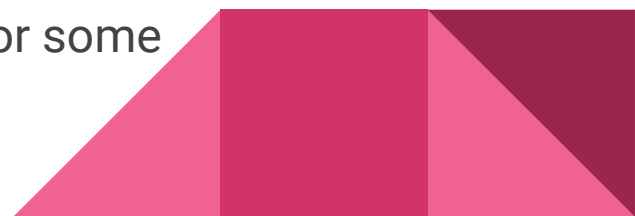Journal articles from web sites that follow indexing guidelines

Indexing guidelines ensure journal isn't partially indexed

Selected conferences in CS/EE (primary publication venues)

Primary publication venues in these fields

Preprint collections from arXiv and NBER

First point of publication for many, final pub venue for some

Scholar Metrics: inclusion

Must have at least 100 articles in 5 years

    Serials need to have continued presence

    Need to draw the line somewhere

Must have at least one citation considering all articles

Includes citations from **all** indexed items (journals, books, preprints,...)
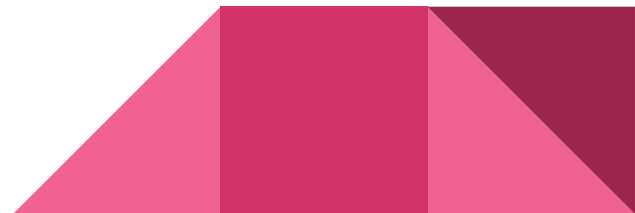
    Not just journals included in Scholar Metrics

Scholar Metrics: coverage

More publication venues than other metrics releases

Includes journals in all major languages

Articles that appear in multiple journals/conferences included for both

Published preprints included for both preprint and journal

Scholar Metrics: presentation

Browse by category (English only) and by language

Journals/conferences in non-ascending order of h5-index

    H5-median used as secondary ordering measure

For each journal, browse list of articles that make up h-index (h-core)

    Allows potential authors to view community of successful papers

    As well as community of successful authors

Presentation - English publications

Categorized lists for English publications

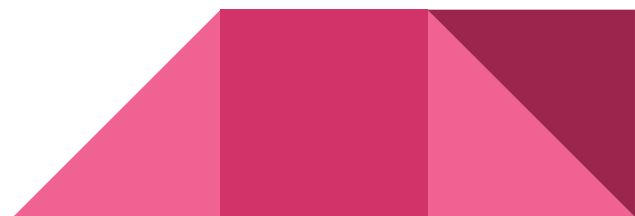   Browse by 8 broad and ~250 specific categories

Journals can appear in multiple categories

    Multi-disciplinary journals, categories with blurred boundaries

Automatically computed clustering & categorization

    Per article categorization

    Clustering across all articles in a journal

Presentation - non-English publications

Language based lists for 11 non-English languages

Chinese, Portuguese, Spanish, German, Russian, French, Japanese,

Korean, Polish, Ukrainian, Indonesian

Languages with largest number of journals

No categorization

Automated categorization harder for fewer journals/category

Presentation - beyond browsable journals

Category browse includes top 20 journals per category by h5-index

Language browse includes top 100 journals per language

Many more journals can be searched using keywords in journal name

Long browse lists hard to navigate (100 items already too many)

H5-index advantages: noise resistant

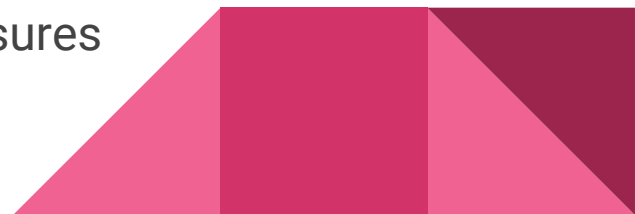Noise source: single very highly cited paper

    Single successful paper does not a community make

    Counting successful papers instead of trying to estimate average

Noise source: extreme diversity in styles, source fidelity, language conventions

    Can result in some variation in citation counts

    Counting papers instead of citations stabilizes measures

H5-index advantages: considers all articles

Avoids sensitivity due to article classification

> News/editorials etc just like uncited articles

> Changes in article classification have no impact on measure
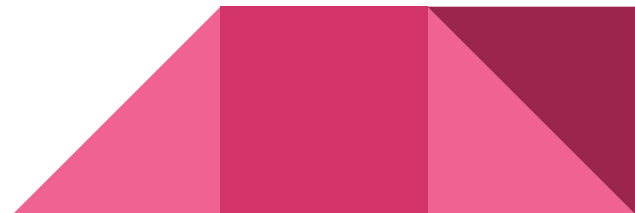
H5-index advantages: quick pick up

New journals can be included in about a year

    In the year after 100 articles are published

    New journals need visibility, older journals are usually known

Conferences have more churn than journals

    New conferences are created frequently

H5-index advantages: stability

Citation counts vary based on specific set of papers

     Approaches that count citations can result in a bouncing measure

     Counting papers tolerates a fair bit of variation in citation counts

Averaging based metrics sensitive to number of papers

     Variation in publication rate or changes in which papers are included

     Focusing on successful papers tolerates change in pub rate
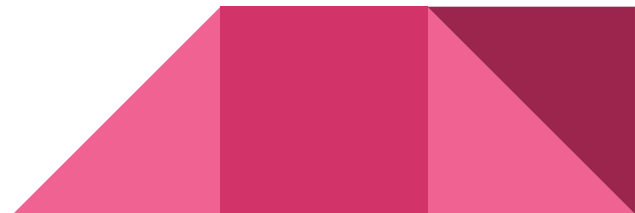
H5-index advantages: easy to follow

No complex determination of citable articles

    All articles are included

No transitive computation

    Transitive computation is hard for a reader to verify

    List of articles in the H-core stand witness to the computation

H-index concern: do larger venues do better?

Adding low cited papers doesn't increase h-index

Adding papers cited just as much doesn't increase h-index

What matters is increasing number of high cited papers

And the number of citations across **all** the high cited papers

Successful papers usually impact direction of fields

Journals with many such papers have large impact on field!

Summary

Metrics that focus on successful papers suitable for online journals

  Fit author aspirations as well as drive usage

H5-index has several advantages as journal metric

  Noise-resistant, tolerant to input variations

  Simple to compute, easy to follow, quick pick up