



The BIP! Toolbox for Scientific Impact Assessment & Applications

Thanasis Vergoulis, Ilias Kanellos, Serafeim Chatzopoulos

IMSI, ATHENA RC, Greece



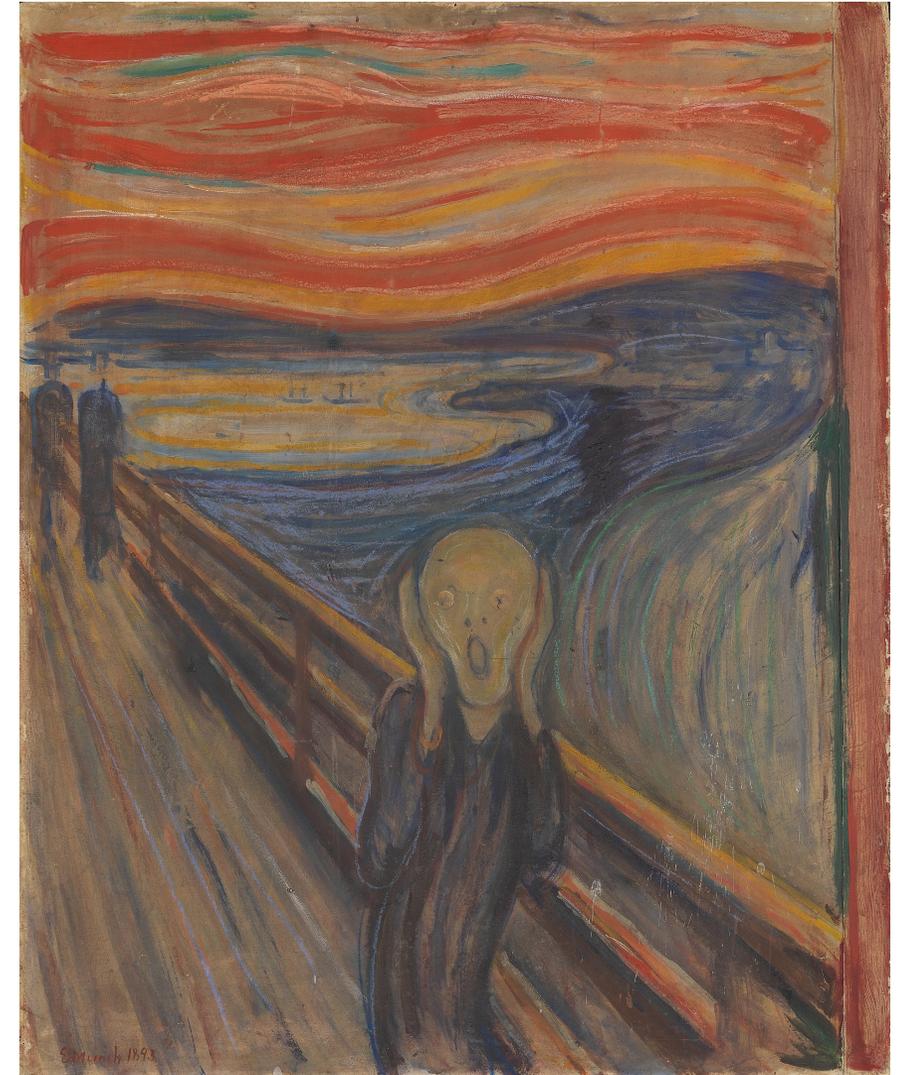
Published research is exponentially increasing

- The ***growth rate*** of the number of published research is ***constantly increasing***.
- Studies suggest that among the vast number of published works many are of ***questionable quality or low impact***.
- Identifying most ***valuable publications*** for any given research topic has become ***tedious & time consuming***.



Why?

- ***Increase in the number of researchers worldwide.***
 - ^20% between 2007-2014*
- ***Publish or Perish***
 - incredible pressure to publish more, especially on young researchers



Edvard Munch, "The scream of nature"
<https://bit.ly/3dcLbXD>

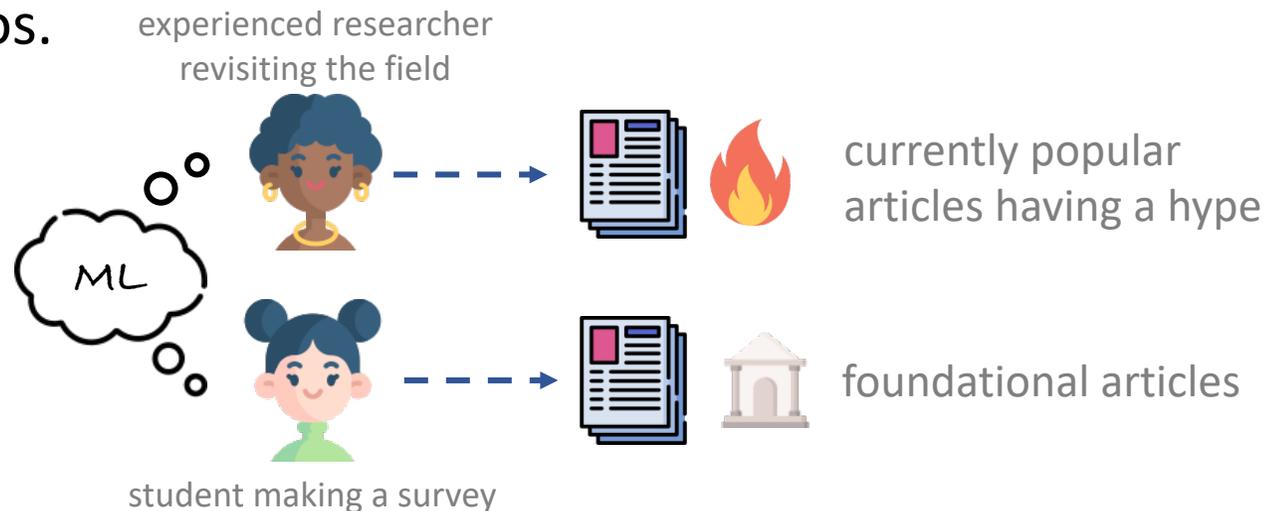
Solution?: impact measures/indicators

- **Quantifying the impact** of publications could facilitate the identification of valuable research.
 - **Open Science initiatives**, having momentum make the calculation of such measures possible.
- Academic search engines **combine keyword-search with a scientific impact measure** (usually citation counts) to rank publications.
 - possible other applications

The screenshot shows a Google Scholar search for 'artificial intelligence'. The search bar at the top contains the text 'artificial intelligence' and a search icon. Below the search bar, it indicates 'Articles' and 'About 3,210,000 results (0.03 sec)'. On the left side, there are filters for 'Any time' (with sub-options: Since 2021, Since 2020, Since 2017, Custom range...), 'Sort by relevance' (with sub-option: Sort by date), and checkboxes for 'include patents' (unchecked) and 'include citations' (checked). At the bottom left, there is a 'Create alert' checkbox. The main results area shows two entries. The first entry is a PDF titled 'Artificial intelligence: a modern approach' by S Russell and P Norvig, published in 2002. It has a citation count of 38141, which is highlighted with a red box. The second entry is a book titled 'An artificial intelligence approach' by R Mitchell, J Michalski, and T Carbonell, published in 2013 by Springer. It has a citation count of 2573. Both entries include links for 'Related articles' and 'All versions'.

Pitfall #1: Scientific impact has various aspects

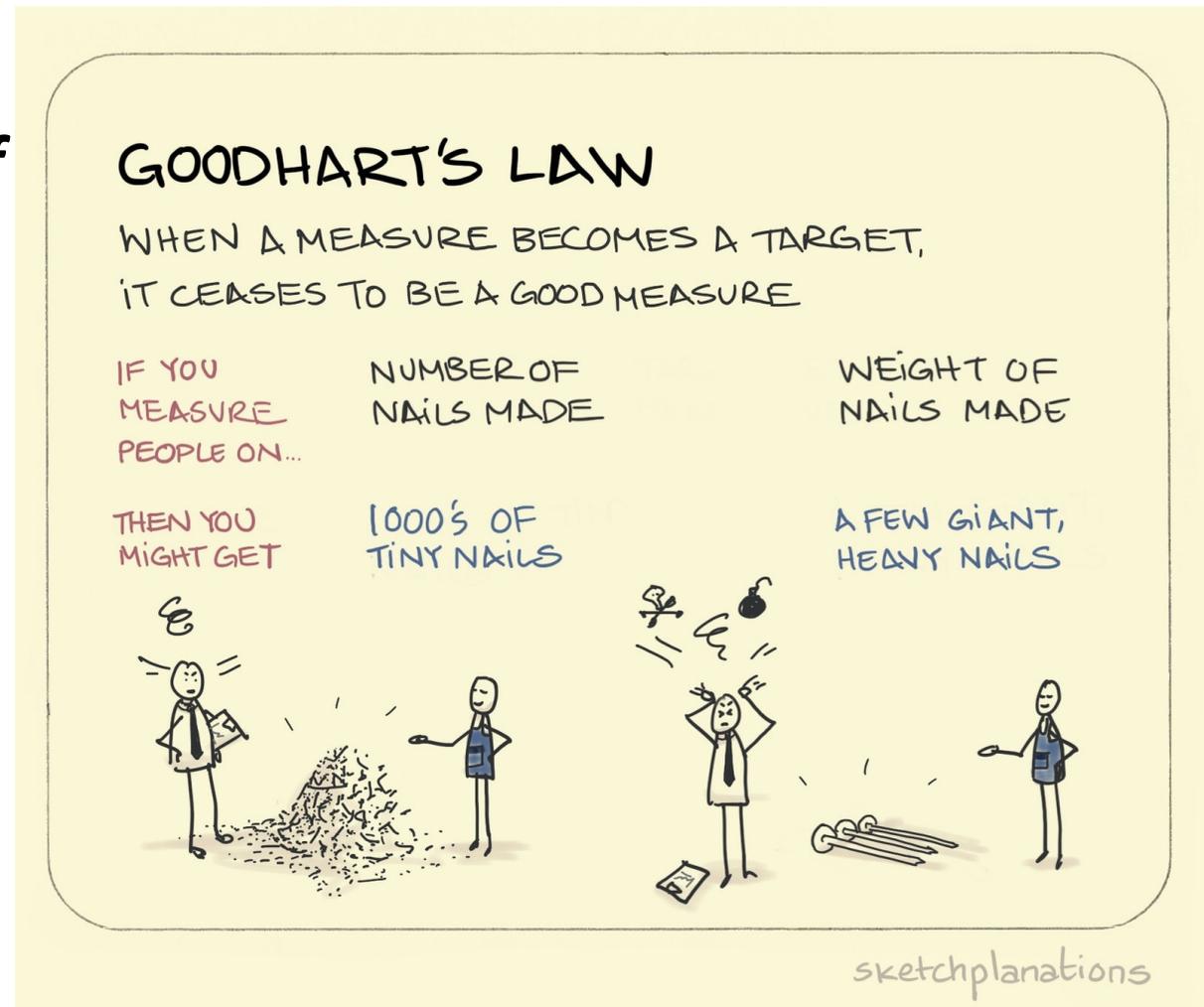
- It is an ***oversimplification to rely only on one impact measure***, like most academic search engines.
 - There are many ***diverse aspects*** of scientific impact, each most appropriate in different scenarios.



- Also there is ***scientific merit***, not only impact...
 - Merit/quality is not completely correlated with impact

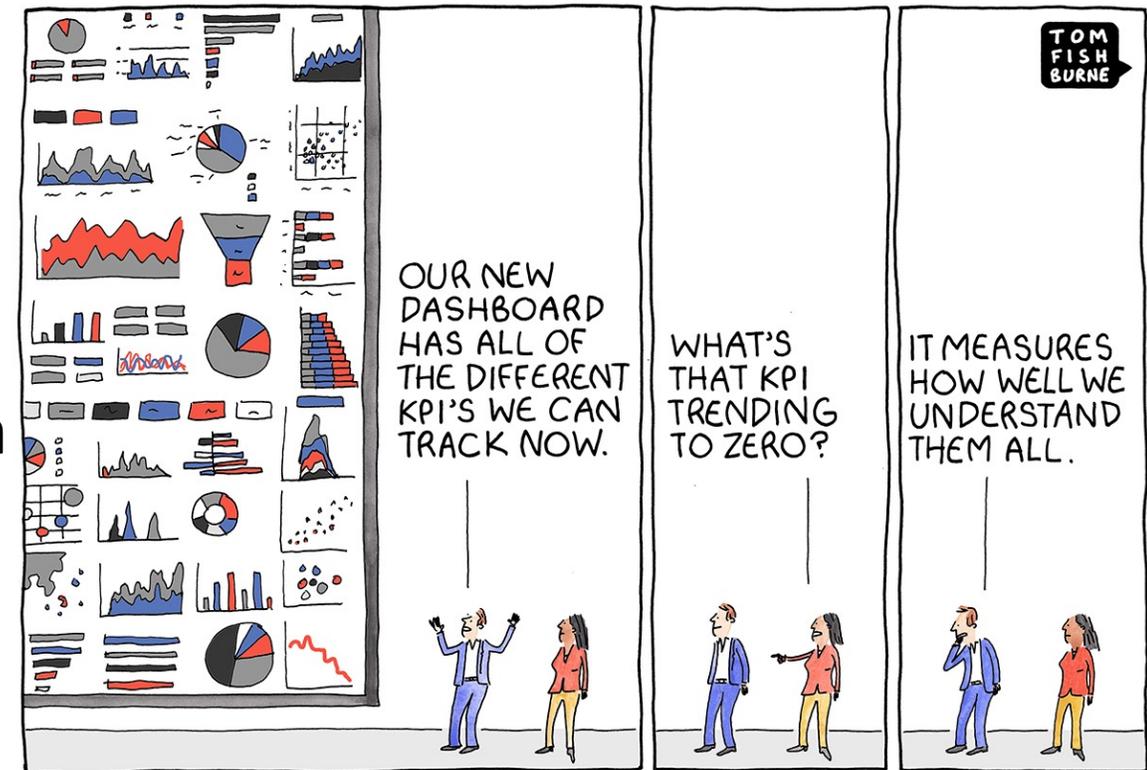
Pitfall #2: Goodhart's/Campell's law

- Scientific impact *should not be examined through a limited set of measures*.
 - Any individual impact measure has *limitations*.
 - More measures capture a *wider range of impact aspects*.
 - **Goodhart's law/Campell's law:** individual measures are vulnerable to attacks & become targets (more measures → increased difficulty for attacks)



Pitfall #3: No proper interpretation

- There is *a multitude of impact measures*.
- In most cases only the measures are provided *without the proper interpretations*, best practices, etc.
- *The landscape is confusing* and often the measures are not properly used.



© marketoonist.com

<https://marketoonist.com/2019/11/kpi-overload.html>

BIP! Toolbox motivation

- To create a set of services & resources to offer a ***multi-dimensional view of publications impact***.
 - PaperRanking library
 - BIP! Finder (search engine, impact-based ranking)
 - BIP! API (open API to get the impact measure scores)
 - BIP! DB (open Dataset)
 - BIP4COVID19 (search engine & open dataset with impact scores for COVID-19 papers)
- We hope that ***3rd parties will build useful services on top of this dataset*** to provide useful services to researchers and other professionals in academia.

120M publications

OpenCitations (Sep 2021)

MAG (Jul 2021)

Crossref (May 2021)

BIP! DB

673

👁 views

93

📄 downloads

BIP4COVID19

149,312

👁 views

21,132

📄 downloads

BIP! Finder



BIP! Finder

Amplifying valuable research

Order by: Popularity Influence Impulse Year
Keyword relevance: Yes No

1448988 results (72450 pages)

« 1 2 3 4 5 »

Click on entries for comparison

Title	Venue	Year	Impact
The psychological impact of quarantine and how to reduce it: rapid review of the evidence ⓘ	context ⓘ The Lancet	2020	🔥 🏛️ 📈 📌
Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2 (...)	context ⓘ International Journal of (...)	2020	🔥 🏛️ 📈 📌
A globally coherent fingerprint of climate change impacts across natural systems ⓘ	context ⓘ Nature	2003	🔥 🏛️ 📈 📌

search, rank & compare papers according to different aspects of their impact

Reproducibility (15)

An introduction to Docker for reproducible research ⓘ

ACM SIGOPS Operating Systems Review - 2015

Read



GUIdock: Using Docker Containers with a Common Graphics User Interface to Address the Repr (...)

PLOS ONE - 2016

Read



Using Docker Containers to Improve Reproducibility in Software and Web Engineering Researc (...)

Lecture Notes in Computer Science - 2016

Read



AlgoRun: a Docker-based packaging system for platform-agnostic implemented algorithms ⓘ

Bioinformatics - 2016

To Read



Using docker containers to improve reproducibility in software engineering research ⓘ

Proceedings of the 38th International Conference on Software (...) - 2016

Read



create and manage your reading list



That's all Folks!

email: vergoulis@athenarc.gr

twitter: [@vergoulis](https://twitter.com/vergoulis) & [@BipFinder](https://twitter.com/BipFinder)

BIP! Dataset at a glance

- **5 citation-based measures** quantifying **3 distinct aspects of scientific impact** for **~117M publications** from various disciplines.
 - Currently: 5th release (version 3)
- Analysis applied on a **citation network** based on a combination of:
 - **OpenCitations' COCI dataset** (Dec-2020)
 - **MAG snapshot** (Nov-2020)
 - **Crossref snapshot** (Jan-2021)
- Openly available (CC-BY license) at Zenodo:
<https://doi.org/10.5281/zenodo.4386934>
- Open API: <https://bip-api.imsi.athenarc.gr/documentation>
- Search engine: <https://bip.imsi.athenarc.gr/> (to be updated soon)



BIP! DB: Currently supported measures

- We ***focused on citation-based measures***.
 - We plan to extend to other types of measures in the future.
- We were ***based on our previous experimental study***:

Ilias Kanellos, Thanasis Vergoulis, Dimitris Sacharidis, Theodore Dalamagas, Yannis Vassiliou:
Impact-Based Ranking of Scientific Publications: A Survey and Experimental Evaluation.
IEEE Trans. Knowl. Data Eng. 33(4): 1567-1584 (2021)
- We tried to include a set of ***measures that capture diverse aspects of scientific impact***:
 - ***Popularity***: a publication's current attention
 - RAM & AttRank
 - ***Influence***: a publication's overall, long-term importance
 - CC & PageRank
 - ***Impulse***: a publication's initial impact during its "incubation phase" (the first years after its publication)
 - "incubation" CC (based on first 3y after publications)

But are they different?

- We measured the ***top-k pairwise correlations*** for the top 1M papers (~1% of the dataset).
- Measures focussing for the ***same impact aspect more correlated to each other***, in general.

influence measures appear to have moderate correlation (larger for larger k vaules, e.g., 10% of articles)

	iCC	CC	PR	AttRank	RAM
iCC	1	0.0985	-0.3468	0.3141	0.3042
CC		1	0.4144	0.4583	0.2774
PR			1	-0.0675	-0.2598
AttRank				1	0.9056
RAM					1

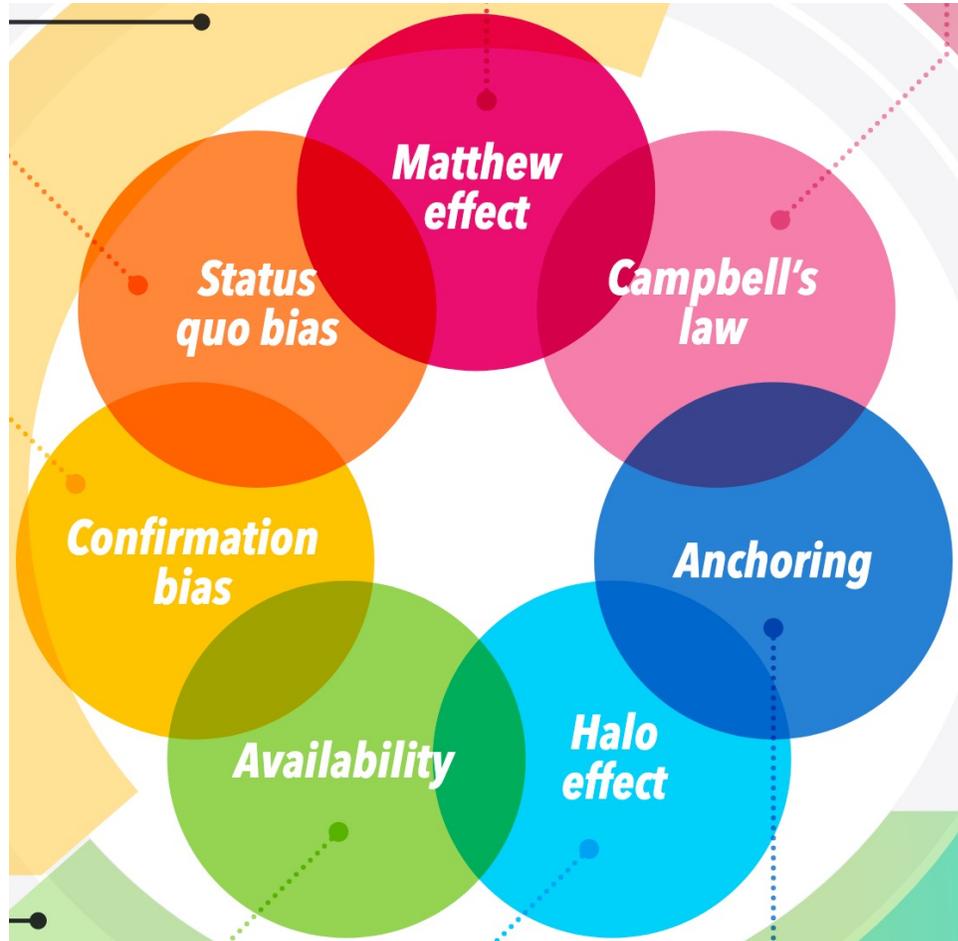
impulse measure seem diverse to other measures

AttRank & RAM (both focussing popularity) are highly correlated

Table 2: Top-1% pairwise correlations of impact measures.

DORA: unintended & cognitive system biases in research assessment

2021



We should *combine qualitative & quantitative information* for research assessment.

We believe that it is important to also have **quantitative information that reflects various, diverse aspects of scientific impact & merit.**

Even more pitfalls...

