

7-1-2010

Metric mad: the future of scholarly evaluation indicators

Ashlea Higgs

Follow this and additional works at: <https://www.researchtrends.com/researchtrends>

Recommended Citation

Higgs, Ashlea (2010) "Metric mad: the future of scholarly evaluation indicators," *Research Trends*: Vol. 1 : Iss. 18 , Article 4.

Available at: <https://www.researchtrends.com/researchtrends/vol1/iss18/4>

This Article is brought to you for free and open access by Research Trends. It has been accepted for inclusion in Research Trends by an authorized editor of Research Trends. For more information, please contact r.herbert@elsevier.com.

Continued from page 6

microscopy – such as Kazu Suenaga and Sumio Iijima at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan, and Alex Zettl at UC Berkeley – were able to translate their expertise directly to graphene. The large-scale growth of graphene using chemical vapor deposition (CVD) was a similar case: groups with experience and apparatus set up for CVD of nanotubes – such as Rodney Ruoff at the University of Texas at Austin – were able to modify the catalyst structure to grow graphene. Surprisingly, it was two scientists

with no background in carbon nanotubes or fullerenes, Kostya Novoselov and Andre Geim, who made the biggest contribution to the field of graphene. This highlights how people from outside the immediate field can make a massive impact.

References:

- [1] Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., Smalley, R.E. (1985) "C60: Buckminsterfullerene", *Nature*, vol. 318, issue 6042, pp. 162-163.
- [2] Iijima, S. (1991) "Helical microtubules of graphitic carbon", *Nature*, vol. 354, issue 6348, pp. 56-58.
- [3] Soldano, C., Mahmood, A., Dujardin, E. (2010) "Production, properties and potential of graphene", *Carbon*, vol. 48, issue 8, pp. 2127-2150.
- [4] Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A. (2004) "Electric field in atomically thin carbon films", *Science*, vol. 306, issue 5696, pp. 666-669.

Reporting back



Metric mad: the future of scholarly evaluation indicators

ASHLEA HIGGS

In mid-December 2009, around 50 science colleagues assembled for what was tipped to be a veritable bibliometric wonderland. Attended by George Hirsch and Henry Small among others, the event offered a practical workshop rather than one-way theoretical presentations.

Jumping on the interdisciplinary bandwagon, the speakers and attendees represented many differing points of view: government vs. academic vs. corporate; evaluator vs. proposer; funding vs. policy vs. scientist; metric theorists vs. practitioners. But while debates were spirited, discussions were collegial and focused on advancing work on new metrics.

Two particular questions occupied participants, to which all discussions of new metrics circled back. Herbert van de Sompel of Los Alamos National Laboratory, the first speaker and one of the event organizers, asked attendees: "What are the qualities which make a metric acceptable to all stakeholders? And how do we move from conception to

acceptance?" The workshop centered on projects investigating or proposing new metrics, including the MESUR project, Eigenfactor, h-bar index, and PLoS ONE's article-level metrics. Many of these new metrics center on usage data.

Usage-based versus article-level metrics

Metrics based on usage data are central to the **MESUR (Metrics from Scholarly Usage of Resources) project**. Johan Bollen from Indiana University, and principal investigator for the MESUR project, presented his findings to date. When comparing

traditional citation-based metrics with usage-based metrics, he observed that usage data are very good indicators of prestige, but that evaluating scholars solely on rate metrics and total citations is "like saying Britney Spears is the most important artist who ever existed because she's sold 50 million records."

In contrast, Peter Binfield of **PLoS ONE** presented the journal's work on article-level metrics. In PLoS ONE, article views, downloads, star ratings, bookmarks and comments join

New usage metrics: recurring themes, fresh challenges

Small beginnings: it took centuries for citation structure to develop; technologies are only now available to make new metrics possible.

Incentives work both ways: people need incentives to adopt new metrics, while metrics incentivize both positive and negative behavior.

Availability of raw data: usage data can be proprietary, fragmented, and not overtly displayed.

Metrics are only part of the answer: peer review continues to play a role.

Continued on page 8

Continued from page 7

the traditional citation counts. There are, however, downsides to article-level metrics like the star-rating system: Peter cautions that it is not yet widely used, and there is a propensity to give articles a five-star rating. The full scale is rarely used, meaning that it can be hard to infer much from these ratings.

Missing from the current metrics available, claims Peter, include those that predict an article's impact from day one; ratings by reviewers, editors and other experts in an article's particular field; mainstream media coverage; publicly available usage metrics that track article downloads, views of abstracts, re-posts of articles online and so on; tracking of "conversations" (comments, forum discussions and so on) outside the original place of publication; and the reputation of metrics among commentators.

Moving with the times

Recognizing that citation analysis has a history hundreds of years in the making, the discussion of new usage indicators has only been possible in the last decade or two. It will take a long time before scholarship catches up with these new (technological) metrics; and we are only just beginning to understand what the impact of these technologies will be.

Will Jorge Hirsch's *h*-bar index take hold with the speed of the *h*-index? Will collaboration between the MESUR and Eigenfactor projects deliver MESUR-able results? Which approaches to network analysis will become mainstream in identifying influence, prestige and trust? When will measuring re-use of data sets become commonplace? Will metrics ever replace peer review? Whatever the answers, we look forward to the next workshop to carry the debate forward.

Useful links:

[Scholarly Evaluation Metrics: Opportunities and Challenges](#)

[Scholars Seek Better Metrics for Assessing Research](#)

[Productivity](#)

[MESUR](#)

[PLoS ONE](#)

[Visualizations on PLoS](#)

The burning questions

Cart before the horse: new usage-based metrics require the collection of new data for future analysis. But what data with what standards and for which metrics?

Variances in usage-tracking systems: without a central repository, how to measure usage across databases for the same article?

Power of simplicity: simple calculations, such as the *h*-index and impact factor, have high adoption rates; will relatively complex, computer-dependent network analyses ever achieve the same rate?

Scholarly vs. public attention: when analyzing usage data of publicly available articles, can scholarly attention be distinguished from general curiosity? Does it need to be?

What's "new": do existing systems for scholarly attention and funding decisions drive attention to the norm, to the detriment of breakthrough research that pushes the boundaries of science?

No single metric: while one metric will never suffice, which set of metrics will serve as a standard group?

Metrics for non-article research output: how can re-purposing mathematical formulas or re-using data sets be tracked?