

Formula predicts research papers' future citations

Mathematical model allows the success of publications — and perhaps scientists — to be forecasted.

Richard Van Noorden

03 October 2013

It sounds like a science administrator's dream — or a scientist's worst nightmare: a formula that predicts how often research papers will be cited. But a team of data scientists now says it could be possible. They report¹ today that a simple model allows reasonably accurate predictions of a paper's future performance on the basis of about five years of its citation history.

"We would like to be able to predict as early as possible, and with relatively stable accuracy, how impactful a particular paper will be in the future," says study co-author Dashun Wang at the IBM Thomas J. Watson Research Center in New York. Others say that the work, published today in *Science*¹, is an interesting advance, but is not yet useful for policy-makers. Even so, it could mark the beginnings of measures that focus on predicting a paper's future influence, rather than evaluating past performance.

The forecasting model relies on picking up clues from how a paper is cited in its early years. Surprisingly, the model does not need to know the subject of the paper, who published it or in which journal. Instead, it assumes that just three basic factors influence how a paper gains citations. The first is, of course, the underlying appeal of its ideas. But another is how immediately it gets cited: if a paper gains an early boost, its visibility makes it more likely to pick up more citations — a well-known 'rich-get-richer' network effect that speeds the paper's approach to peak influence². A third is that novelty fades; eventually, a paper's citation rate approaches zero.

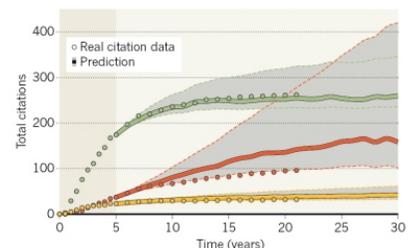
Ultimate impact

Wang, working with Albert-László Barabási, a network theorist at Northeastern University in Boston, Massachusetts, and with Chaoming Song, a physicist at the University of Miami in Florida, built a model suggesting that if relative differences in these factors were mathematically corrected for, all papers would follow the same citation pattern over time. For any research paper, it is then a matter of finding out which relative values (for underlying appeal, rate of initial growth and decay rate) best adjust the observed citation pattern to the universal curve. And with those values in hand, the model can predict a range for the paper's future performance and its probable lifetime citations — or 'ultimate impact', as the team calls it.

The researchers tested out their model on physics papers from the 1960s. They made predictions on the basis of five years' citation data, and found that 25 years later, 93.5% of papers fell within their predicted citation range. In fact, says Wang, forecasts can be made for many papers on the basis of less than five years' data, because their citations peak at around two years and then die out. The model also works on papers from the 1990s and 2000s, he says.

PREDICTING THE FUTURE

A model forecasts how many citations a research paper will pick up, on the basis of five years of citation data.



Data for three papers published in *Physical Review Letters* in 1990.

DATA FROM DASHUN WANG / SCIENCE

Related stories

- Halt the avalanche of performance metrics
- Are scientific reputations boosted artificially?
- Metrics: A profusion of measures

More related stories



Nature special: Metrics

In some cases, however, there was a lot of uncertainty, so the range of future prediction was very wide (see 'Predicting the future' above). And 6.5% of papers defied the model entirely: these were ones that did not stand out in their first five years, but gained a second wind and became influential later on.

Anthony van Raan and Ludo Waltman, who study science citation networks and mapping at Leiden University in the Netherlands, told *Nature* that the model was elegant and the paper important. But policy-makers should not get excited just yet, says van Raan. "Prediction of citation impact five years after a publication's appearance is of little use in a policy context." And he cautions that even if a range of lifetime citation figures can be predicted, administrators should remember that citations inevitably differ between fields; for example, biologists cite each other more than physicists.

Increasing complexity

Wang says that he will improve the model by incorporating more complex elements — such as the paper's topic, or where it was published. "The focus here is on the minimal factors needed. The surprising thing to me is we can achieve this level of predictability just by looking at the citations over time," he says.

The model is also applicable to collections of papers — for example, all the papers published in a given journal, by one institute or by a particular scientist. The last prospect is intriguing, because existing metrics by which scientists are judged, such as the *h*-index, have little capacity to predict future performance, says Wang. Although his model can predict how a scientist's past papers will fare in later years, that does not imply that the scientist's future papers will have similar impact. Even so, says Wang, just finding out how a scientist's future impact relates to past impact would be useful, because it would allow one to quantify, on the basis of citations, "to what extent an individual scientist's career is predictable". Wang now hopes to build a website that would produce citation forecasts for any research paper.

If administrators did use metrics to predict future impact, it could change how science is done, says James Evans, a sociologist at the University of Chicago in Illinois who wrote an article³ on the future of research to accompany Wang's paper in *Science*. Scientific discovery might move faster than it already does, he says. But, he warns, "knowing only the momentum of an article's reception could act as a self-fulfilling prophecy" — where everyone would cluster around hot papers and ditch research areas that might later have proven fruitful.

Nature doi:10.1038/nature.2013.13881

References

1. Wang, D., Song, C. & Barabási, A-L. *Science* **342**, 127–132 (2013).

Show context

2. Barabási, A-L. & Albert, R. *Science* **286**, 509–512 (1999).

Show context

Article PubMed ISI

3. Evans, J. A. *Science* **342**, 44–45 (2013).

Show context

Article

Related stories and links

From nature.com

- **Halt the avalanche of performance metrics**
14 August 2013
- **Are scientific reputations boosted artificially?**

06 May 2011

- **Metrics: A profusion of measures**

16 June 2010

- **Nature special: Metrics**

From elsewhere

- **Dashun Wang**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments

[Subscribe to comments](#)

There are currently no comments.

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2013 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER