

New model reveals forgotten influencers and 'sleeping beauties' of science

19 March 2018, by Rob Mitchum



A new paper led by the Knowledge Lab gives a different measure of influencers in scholarship. Credit: University of Chicago

For centuries, scientists and scholars have measured the influence of individuals and discoveries through citations, a crude statistic subject to biases, politics and other distortions. A new paper led by the Knowledge Lab at the University of Chicago describes a different way to keep score in science—a more direct measure of how influential ideas ripple out across scholarship and culture.

The computational model throws the spotlight onto work that changed the path of science but has remained underappreciated. The same approach also can be adapted to trace influence in other areas where no culture of [citation](#) exists, such as literature or music, said the authors of the paper published last week in *Proceedings of the National Academy of Sciences*.

"We're measuring how much scientists' and scholars' writings influence discussion of ideas in the future," said James Evans, director of Knowledge Lab and professor of sociology at UChicago. "Influence is a politicized process; those who get the influence, get the credit, and those

who get the credit get the capital to do the next big thing. This is the first time we have a tightened ability to identify influence, and also to diagnose social and strategic [influences](#) on citing behavior."

The new paper complements previous Knowledge Lab research using computational and machine learning approaches on massive collections of text, grants, reviews, citations and scientific data to study how discoveries form, evolve and become widely accepted. Their work was recently featured in a review in the journal *Science*, co-authored by Evans, that described how data-driven methods have deepened understanding of the scientific process and offered new insights on how to more efficiently make future important discoveries.

Going beyond citations

In theory, references in an academic paper enable authors to credit their predecessors, the researchers and work upon which they built their new discovery. But in practice, citations are chosen for many reasons—authors are more likely to cite themselves, powerful colleagues in their field and researchers at prestigious institutions, and are often biased towards citing more recent or already highly-cited articles.

Despite these imperfections, many computational studies of scientific influence have used the citation record as a useful proxy. The new study, led by former Knowledge Lab postdoctoral researcher Aaron Gerow, demonstrates a novel, deeper approach, using both the full text of articles and external information such as author identity, affiliation and journal reputation.

Using a computational method known as topic modeling—invented by co-author David Blei of Columbia University—the model tracks "discursive influence," or recurring words and phrases through historical texts that measure how scholars actually talk about a field, instead of just their attributions.

To determine a given paper's influence, the researchers could statistically remove it from history and see how scientific discourse would have unfolded without its contribution.

"We can not only find out how topics changed over time but can actually simulate the future without a given document from the past, and look at how discourse moving forward was different with and without a given document," said Gerow, now an assistant professor at Goldsmiths, University of London. "Citations are one kind of impact, and discursive influence is a different kind. Neither one is the complete story, but they work together to give a better picture of what's influencing science."

Training the model on massive text collections from computational linguistics, physics, and across science and scholarship (JSTOR), the authors quantify various biases and discern distinct patterns of influence. Scientists who persistently published in a single field were more likely to be "canonized" in a way that compelled others to cite them disproportionate to their papers' discursive contributions. On the other hand, discoveries that crossed disciplinary boundaries were more likely to have outsized discursive impact but fewer citations, likely because the "owner" of the idea and her allies remain socially and institutionally distant from the citing author.

Sleeping beauties and unknown influencers

One interesting subcategory of paper the model detected is known as "sleeping beauties," or papers that went relatively unacknowledged for years or even decades before experiencing a late burst of citations. For example, a 1947 paper on graphene remained obscure and forgotten until the 1990s with a resurgence of research interest in the material and an eventual Nobel Prize.

"Papers have a news cycle, when lots of people chat about them and cite them, and then they're no longer new news," Evans said. "Our model shows that some papers have much more influence than citations will typically demonstrate, such as these 'sleeping beauties,' which didn't have much influence early but come to be appreciated and important later."

The same model can also be used to trace influence in other areas, such as literature and music, the authors said. Text from poems or song lyrics, and even extra-textual characteristics such as stanza structure or chord progressions, could feed into the model to find under-recognized influencers and map the spread of new concepts and innovations.

"There's an enormous amount of literary culture which ends up influencing all kinds of things, but which simply does not have a technology of reference similar to citations," Evans said. "Though we developed and validated on this model on scientific text, now we can use it for anything and everything, especially cases where there are no traces of influence but patterns in the content itself. It's like trending on Twitter, but where everything is Twitter. That is what's most exciting to me."

More information: Aaron Gerow et al. Measuring discursive influence across scholarship, *Proceedings of the National Academy of Sciences* (2018). [DOI: 10.1073/pnas.1719792115](https://doi.org/10.1073/pnas.1719792115)

Provided by University of Chicago

APA citation: New model reveals forgotten influencers and 'sleeping beauties' of science (2018, March 19) retrieved 24 October 2020 from <https://phys.org/news/2018-03-reveals-forgotten-beauties-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.