

SCIENTIFIC PUBLISHING

Google Scholar Wins Raves—But Can It Be Trusted?

Over the past year, Jonathan Eisen's reading habits have changed dramatically. For most of the past 2 decades, he has kept up with scientific literature primarily by combing PubMed, the vast trove of online biology abstracts. But these days Eisen, an evolutionary biologist at the University of California, Davis, discovers research relevant to his own work without even looking for it.

The insightful librarian helping keep Eisen up to speed is Google Scholar, a free academic search service maintained by the California-based company. Google Scholar has been studying Eisen closely. It keeps track not only of his own 300 papers and the key words within them—Archaea, *Plasmodium*, phylogenomics—but also the 38,000 citations to his work in published papers, preprint abstracts, books, and even conference posters. Like a scientific version of the Netflix movie recommendation engine, Google Scholar scours the Internet, scoring all scientific documents for their predicted interest to Eisen, and then sends him a weekly e-mail of recommended reading.

Eisen is part of growing crowd of converts. “Google Scholar is having a great impact on the research-seeking behavior of researchers,” says Nicolás Robinson-García, a bibliometric researcher at the University of Granada in Spain. Robinson-García claims that Google Scholar's compendium of articles is at least as comprehensive as the leading commercial academic search databases—Thomson Reuters' Web of Science and Elsevier's Scopus—and for many disciplines in the social sciences and humanities, even better. And by all accounts, it is gobbling up market share. “Google is the dominant referral source for online journal articles, far surpassing the amount of traffic driven by other discovery tools,” says David Crotty, senior editor at Oxford University Press in New York City.

But researchers aren't just using Google Scholar as a search engine. Its algorithm provides citation metrics that quantify the impact of their own published work, and these numbers are becoming part of a standard scientific CV. This byproduct of

Google Scholar has sparked a new concern: Because it includes sources from across the Internet—not only vetted journals—and has no human curators, Google Scholar's citation metric can be easily gamed.

Robinson-García is part of a team that demonstrated that vulnerability by placing six fake papers with long lists of citations to their own work on a webpage within the University of Granada Internet domain. Google Scholar's algorithm dutifully tallied them as real citations and in a matter of weeks, their Google Scholar citation scores rose significantly. The team's findings appeared

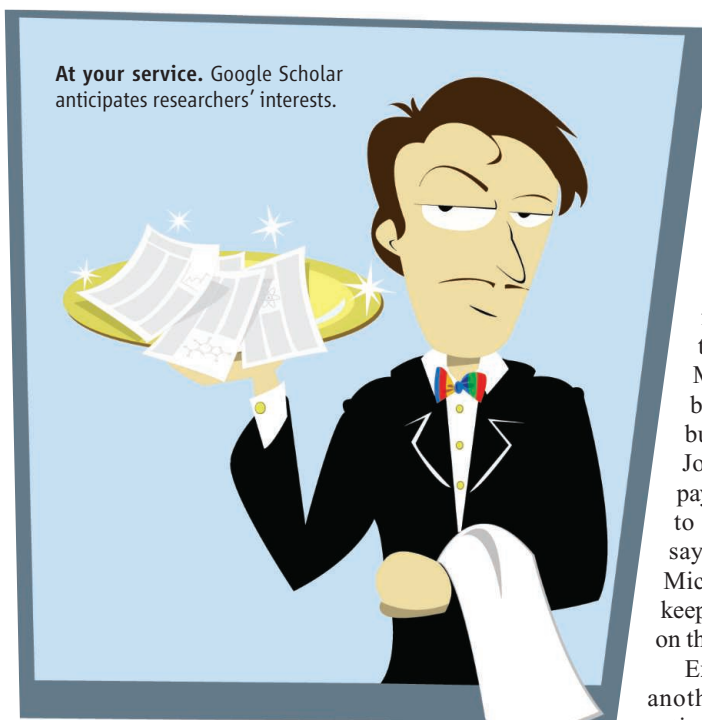
its development, labels efforts to skew the citation metric as “spam.” So far, he says, “the level of spam in scholarly articles remains low.” Acharya chalks this up to the “large penalty that would come with being caught” rigging the system in scholarly communities. But if social norms fail to keep academic cheaters in check, he says, “we can and will adjust the level of spam-handling.” However, he says, Google has no intention of revealing its algorithm, in part because it is tied up with the company's core search engine.

Google Scholar's ascendance may soon be challenged. “Microsoft is still working in this area too,” says Anne-Wil Harzing, the creator of Publish or Perish, a citation analysis tool that uses data from Google Scholar. A service called Microsoft Academic Search is expanding rapidly to cover all academic domains. Meanwhile, Thomson Reuters and Elsevier have a strong financial incentive to keep their services competitive. Many universities are bound by nondisclosure agreements, but Cornell University librarian John Saylor says his institution pays \$155,000 per year for access to Web of Science. As Harzing says, “having both Google and Microsoft in this field will surely keep Thomson Reuters and Elsevier on their toes!”

Experts say the competition has another benefit: giving scientists options, so that they don't come to depend on just one service, such as Google Scholar, that may be vulnerable to corporate downsizing. There are “persistent rumors that Google is de-emphasizing or even dismantling the Scholar team,” says John Sack, founding director of HighWire, an online publishing platform for more than 1300 journals, including *Science*.

Although Google Scholar generates no direct income, Acharya is upbeat about its future. While he declines to reveal usage figures, he claims that the number of users is growing worldwide, particularly in China. And the Google Scholar team is expanding, not contracting, he says. “Rumors of our demise are greatly exaggerated.”

—JOHN BOHANNON



online on 11 November in the *Journal of the American Society for Information Science and Technology*.

If Google Scholar were to provide a breakdown of the sources it tapped for its citation metric, Robinson-García says, “our fraud could have been easily detected.” That lack of transparency is a deal breaker for some bibliometric researchers, including Rodrigo Costas Comesaña of Leiden University in the Netherlands; he calls Google Scholar “an unmanageable tool” for citation analysis.

Google counters that critics are overstating the problem. Anurag Acharya, a co-founder of Google Scholar who leads