

SCHEMES SELLING FAKE CITATIONS ALARM SCIENTISTS

The ways in which researchers can artificially inflate their reference counts are growing.

By Dalmeeth Singh Chawla

Research-integrity watchers are concerned about the growing number of ways in which scientists can fake or manipulate the citation counts of their studies. In recent months, increasingly bold practices have surfaced. One approach was revealed through a sting operation in which a group of researchers bought 50 citations to pad out the Google Scholar profile of a fake scientist they had created.

The scientists bought the citations for US\$300 from a firm that seems to sell bogus citations in bulk. This confirms the existence of a black market for faked references that research-integrity sleuths have long speculated about, says the team.

“We started to notice several Google Scholar profiles with questionable citation trends,” says Yasir Zaki, a computer scientist at New York University (NYU) Abu Dhabi, whose team described its sting operation in a February preprint (H. Ibrahim *et al.* Preprint at arXiv <https://doi.org/ndh9; 2024>). “When a manuscript acquires hundreds of citations within days of publication, or when a scientist has an abrupt and large rise in citations, you know something is wrong.”

These practices are troublesome because many aspects of a researcher’s career depend on how many references their papers garner. Many institutions use citation counts to evaluate scientists, and citation numbers inform metrics such as the *h*-index, which aims to measure scholars’ productivity and the impact of their studies.

References for sale

Research-integrity watchers had already suspected that citations were being sold at paper mills, services that churn out low-quality studies and sell authorship slots on accepted papers, says Cyril Labbé, a computer scientist at Grenoble Alpes University in France. “Paper mills have the ability to insert citations into papers that they are selling,” he says.

In November 2023, analytics firm Clarivate in Philadelphia, Pennsylvania, excluded more than 1,000 people from its annual list of highly cited researchers because of suspicions of citation gaming and ‘hyper-publishing’.

In their sting operation, Zaki and his

colleagues created a Google Scholar profile for a fictional scientist and uploaded 20 made-up studies that were created using artificial intelligence.

The team then approached a company, which it had found while analysing suspicious citations linked to one of the authors in their data set, that seemed to be selling citations to Google Scholar profiles. The study authors contacted the firm by e-mail and later communicated through the instant-messaging service WhatsApp. The company offered 50 citations for \$300 or 100 citations for \$500. The authors opted for the first option and 40 days later 50 citations from studies in 22 journals – 14 of which are indexed by the scholarly database Scopus – were added to the fictional researcher’s Google Scholar profile.

The team didn’t share the company’s name with *Nature*, citing concerns that revealing it could draw attention to its website, or the fake Google Scholar profile they created, and that this might reveal the identities of the authors of the studies that planted the fake citations.

“Citations can easily be manipulated by creating fake preprints and through paid services.”

Asked by *Nature* whether Google Scholar is aware that faked profiles can be created on its site, Anurag Acharya, a distinguished engineer at Google, said: “While academic misbehaviour is possible, it’s rare because all aspects are visible – articles indexed, articles included by an author on their profile, articles citing an author, where the citing articles are hosted and so on. Anyone in the world can call you on it.”

In another demonstration of citation manipulation, last month researchers created a fake Google Scholar profile for a cat called Larry, listing a dozen fake papers with Larry as the sole author. The researchers posted a dozen other nonsensical studies on the academic social-networking site ResearchGate that cited Larry’s papers. After Larry’s identity was revealed, Google Scholar removed the cat’s studies, those citing Larry’s papers, and the accumulated citations. ResearchGate has also

removed the bogus studies citing Larry’s work.

Zaki and his colleagues’ sting operation was born out of a broader effort to assess the scale of the fake-citation problem. The authors used software to examine about 1.6 million Google Scholar profiles that had at least 10 publications. They searched for profiles with more than 200 citations and instances in which researchers’ citations increased by 10 times or more each year or when the rise represented a jump of at least 25% of individuals’ total citation count. The team found 1,016 such profiles.

Zaki says that many citations to the papers on those profiles are from preprint articles that haven’t been peer reviewed and that they are typically listed in the bibliographies of papers, but not cited in the main body of the manuscripts.

“Citations can easily be manipulated by creating fake preprints and through paid services,” says study co-author Talal Rahwan, a computer scientist at NYU Abu Dhabi.

The authors also surveyed 574 researchers working at the 10 highest-ranked universities in the world. They found that of those universities that consider citation counts when evaluating scientists, more than 60% obtain these data from Google Scholar.

Fishy patterns

Labbé isn’t convinced by the survey’s claim that Google Scholar is widely used to obtain researchers’ citation metrics. Allegations of citation manipulation on Google Scholar have surfaced in the past, he says, and academics have long suspected that there are vendors offering this sort of service. But the sting operation to reveal a citation seller is the first of its kind, he says.

Guillaume Cabanac, a computer scientist at the University of Toulouse in France who has created a tool that flags fabricated papers that contain odd turns of phrase added to circumvent plagiarism-detection software, says that many studies are cropping up with citations to work that has nothing to do with the topic of the study.

Labbé’s team is building a tool that automatically flags fishy citation patterns that might point to manipulation.

To help with that, Zaki’s team proposes a metric called the citation-concentration index, designed to detect cases in which a scientist receives many citations from only a few sources. Such activity is often a sign of a ‘citation ring’, in which scientists agree to cite one another to inflate each other’s metrics. “Suspicious ones tend to have massive citations stemming from just a few sources,” Rahwan says.

One fear among integrity sleuths is that fraudsters will conceive subtler practices to avoid being found out. For instance, one way to avoid being detected by the citation-concentration index, Labbé notes, is to buy a few citations at a time, not in bulk.