

Editorial

# Misconduct, Mishaps, and Misranking in Bibliometric Databases: Inflating the Production and Impact of Scientists

Leandros Maglaras <sup>1,\*</sup>  and Dimitrios Katsaros <sup>2</sup><sup>1</sup> School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, UK<sup>2</sup> Faculty of Engineering, University of Thessaly, 382 21 Volos, Greece; dkatsar@inf.uth.gr

\* Correspondence: leandros.maglaras2@dmu.ac.uk

## 1. Introduction

Today, academics and researchers constantly strive to achieve more in their respective fields. Their achievements are measured mainly by how many publications they have within publication venues and their work's recognition (impact), which is usually determined through its citations, subsequently affecting how funding and awards [1] are obtained. To assess the importance academics place on citations when evaluating scientists for recruitment or promotion, the authors of [2] surveyed faculty members from the top 10 ranked universities globally. Their findings indicate that the majority of faculty members take citation counts into account when assessing candidates, which is reflected at a local and national level [3].

The availability of huge curated bibliographic databases such as Elsevier Scopus and Web of Science (WoS) [4] over the past twenty years has led decision-makers involved in promotions, funding, and strategic direction to increasingly request data related to individual studies or scholars (such as scientific articles, PhD students, postdoctoral researchers, and faculty members) as well as groups of individuals and articles (such as journals, universities, institutions, and companies) to support their decisions.

Publication practices in the fields of social sciences and humanities differ from those used for most natural science publications. Consequently, their research output is often inadequately represented in the aforementioned journal-based databases typically used for bibliometric analysis. This issue is particularly pronounced for non-English journals, which are notably underrepresented, as well as for conference papers, books, and edited volumes [5]. An alternative to the more traditional journal-based systems of WoS and Scopus is Google Scholar (hereafter referred to as GS), which is one of the most comprehensive databases currently available. Several works, e.g., [6], have analyzed the relative coverage between Google Scholar and Scopus.

As soon as scientists realized that a significant proportion of their evaluation was based on these purely quantitative methods, some started to take advantage of the system. At first, the prevalence of *plagiarism* was sparse. However, many members of the academic community soon began consistently striving to optimize their performance through two key approaches: (a) increasing the number of papers they have authored and (b) increasing their impact, i.e., the number of citations received by these papers. While it is of course acceptable for a scientist to increase their productivity and the quality of their research impact to attract more citations, several malpractices [7] started making their appearance in the academic landscape. Some malpractices used to optimize authorship include buying authorship [8] and generating large authorship lists by merging and splitting articles. Some of the malpractices used to optimize impact include the use of excessive self-citations, citation circles, and coercive citations [9], as well as uploading fake documents, editorial grouping, and using Generative AI tools [10].

Most of these malpractices are easily achievable in Google Scholar since it is editable by the end user, but some, such as self-citations, citation circles, and coercive citations, are



**Citation:** Maglaras, L.; Katsaros, D. Misconduct, Mishaps, and Misranking in Bibliometric Databases: Inflating the Production and Impact of Scientists. *Computers* **2024**, *13*, 287. <https://doi.org/10.3390/computers13110287>

Received: 4 November 2024

Accepted: 5 November 2024

Published: 7 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

also a problem for curated bibliographic databases. Additionally, quality control issues in Google Scholar exacerbate the situation. In the remainder of this editorial, we will briefly describe the mechanisms behind these malpractices and provide some ideas for reducing the problem.

## 2. Manipulation Tactics

### 2.1. Inflating the Number of Published Articles

In order to achieve an increase in paper production, the easiest way was/is to publish in *predatory* publishers. Many of these have been identified in the academic market. To put it simply, such publishers publish whatever article is submitted to them, most of the time without adopting a reviewing procedure or having a very “light” one; of course, they charge a rather high publication fee for their service. This practice is so extended and popular that the popular Greek newspaper “KATHIMERINI” recently published a small-scale study (<https://www.kathimerini.gr/society/563268901/akadimaikoi-pontoime-pliromenes-dimosieyseis/>, accessed on 1 November 2024), which revealed that in a peripheral Greek university—which was recently merged with a former Higher Technological Educational Institute in its region—39% of publications by its faculty members appeared in predatory periodicals.

A popular way to “artificially” increase paper production is to *buy authorship* [8]. Often, in between a paper’s submission and its final acceptance, the number of coauthors increases significantly and without a solid reason, essentially meaning that the new coauthors did not contribute significantly, e.g., in the revision of the initial submission. This practice is not very effective anymore since both good- and premium-quality conferences and journals apply very strict rules when it comes to adding (or removing) coauthors from the initial submission, but it is still an issue.

Another unethical way to increase the count of published papers is so-called *gift authorship*, which concerns including people as authors who have not contributed substantially to the research. A common reason for gift authorship is due to reciprocity or deference. Both gift and paid authorship result in articles with relatively large authorship lists.

A similar problem concerns articles with *large authorship lists* with over 40 or 50 authors. In this case, uncertainty arises from questions about the authors’ contributions to these papers. The number of multi-authored articles has risen over the last decade, with many including individuals who made only minor contributions [11]. Sometimes, the author list is so extensive that it is comparable to or even bigger than the abstract’s length. For instance, a report on the Large Hadron Collider in the physical sciences featured nearly 3000 authors, while a clinical trial in The New England Journal of Medicine listed 974 authors [11,12].

### 2.2. Inflating the Number of Received Citations

For a long time, two popular techniques for increasing citations have been used, namely *self-citations* and *citation circles*. The “problem of citation circles” refers to a phenomenon where a small group of researchers or academic publications excessively cite each other’s work, creating a closed loop of citations. This issue raises concerns about the integrity and fairness of the academic citation system. It can distort the impact and quality of research and lead to biased evaluations in areas like academic ranking, funding decisions, and peer reviews. Self-citation is a special case where authors cite their own work [13]. In addition to the fact that self-citations might contribute to an author’s scientific progress, as explained in [14], they also act as a citation-boosting technique when used excessively.

A second malpractice is known as *coercive citations*, wherein an editor or referee of a journal/conference asks (or in fact forces) an author to add citations to an article before the journal will agree to publish it. Usually, this is carried out to inflate the journal’s Impact Factor, but other times it is purely for personal “profit”. For instance, the second author of the present article, along with his former PhD advisor, discovered that a member of two prestigious journals’ editorial boards was published by a “giant” publisher. Many of the articles handled by this publisher shared a strange characteristic: they all extensively cited

articles written by this particular editorial board member, even though the cited article had little or no relevance at all to the citing article. The References list of these articles consisted of between 20% and 40% citations to articles written by this editorial board member, with the author list of the cited article abbreviated (i.e., hidden) behind “et al.” so that the name of the editorial board member did not appear in the citations.

Another malpractice in the field is uploading *fake documents* that are automatically indexed by Google [15]. In [16], the authors conducted an experiment in order to showcase this problem. Initially, they created six documents linked to a fake author and uploaded them to a researcher’s webpage within the University of Granada’s domain. The outcome was an increased number of citations (774) for 129 papers (an average of 6 citations per paper), boosting both the authors’ and journals’ *h*-index.

Many researchers focus on the problem of scholars *merging* articles to increase their *h*-index [17] or other indices, such as the *g*-index and the *i10*-index [18]. A similar technique is *splitting* articles [19] to increase the *h*-index of an author.

Another method being used to manipulate GS is reported in [20], namely *citation bazaar*. This 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.

Recently, we have identified a very problematic and rather unethical malpractice: *editorial merging*, where editors group editorials and conference proceedings under their names while their actual contribution only concerns the editorial or the management of the conference. Incidentally, we have detected that many highly-ranked professors have followed this trend. Such a technique increases their position within their institution, thus justifying their superiority [17]. There are many prestigious academics whose Google Scholar profile has over 10,000 citations while their Scopus profile has under 3000 or even 1000 citations. Furthermore, in a recent study [20], a decline in citation counts due to Scopus’ more selective indexing was expected; however, suspicious authors experience an average citation drop of 96% on Scopus, compared to a 43% average drop for normal authors. This strongly indicates their profile being extensively manipulated by merging others’ articles into their profile under the assumption that when editing a book or conference proceedings, you own the intellectual property and work of the contributors.

This aforementioned “grouping” can sometimes happen accidentally, e.g., due to GS’s quality control issues; for instance, the famous 1986 backpropagation paper by Rumelhart et al. is shown in GS with 55,000 citations, but 28,000 of them cite the whole book in which this specific article appeared.

In the following Table 1, we present some examples of GS misuse without revealing the real names of those academics.

**Table 1.** Examples of GS misuse.

ID	Google Scholar	SCOPUS	Drop	Method
Author 1	72,000	130	99%	Article Inclusion
Author 2	480,000	3500	99%	Name Merging
Author 3	19,000	3500	81%	Editorial Merging

### 3. Discussion

In this editorial, we have briefly discussed common methods of bibliometric misconduct with a focus on GS manipulation attempts. These include citation circles, fake documents, the merging of documents and profiles, and more. The fight against publication/citation misconduct is continuous, and although AI tools may worsen the situation through the massive production of AI-generated articles [21], they can also be used as a tool against misconduct [22].

Here, we list rules of thumb for identifying a fake Google Scholar account:

- Lack of verified email domain: Legitimate profiles often have a “Verified email at [institution]” label next to their name. Fake profiles usually do not have this.
- Unrelated publications: Check if the listed publications seem random or unrelated to the claimed field of expertise. Fake profiles often have unrelated or suspicious titles.
- Suspicious citation patterns: If a profile has an unusually high number of citations or self-citations (i.e., citing their own papers excessively), this could be a red flag.
- Inconsistent or incorrect details: Look for inconsistencies in the bio, institution, or profile picture. Often, fake profiles copy famous researchers’ publications or use generic images.
- No academic history or institution links: Fake profiles may not be linked to any institution’s website and have minimal information about the researcher’s background.
- Recent account creation with excessive citations: If an account is newly created but shows an unusually high number of citations in a short time, it could be suspicious.
- Publication list with too many asterisks after article entries: The asterisks represent a merging of articles; in cases of editorial misconduct, they can be used to extensively increase the citation count of an author.

Some recommendations for addressing the issues raised in this editorial are given below.

- **Transparent Citation Practices:** Journals and institutions can encourage authors to provide some reasoning behind their citations and avoid unnecessary self-citations or loose citations. Of course, this is rather difficult to apply to all publications. Editors and reviewers can play a significant role in safeguarding the integrity of publications. Moreover, ensuring a diverse range of reviewers are included in peer-review processes can help minimize the bias caused by citation circles. Additionally, concepts such as *coterminal citations* [14] and their use in indices can be used for tracking phenomena such as citation circles or excessive self-citing.
- **Metrics Beyond Citations:** Academic institutions can adopt more holistic measures of research impact. These should go beyond citation counts and could include other metrics such as societal impact or collaborations outside an author’s discipline. Many institutions are already taking action in this direction.
- **Algorithmic Monitoring:** Some organizations have developed tools to monitor and detect suspicious citation patterns, alerting editors or institutions to potential citation manipulation.
- **Correlation between citation metrics:** Comparing and correlating GS with WoS, Scopus, or other databases can provide a more accurate reflection of a scholar’s real merit.
- **Use of databases that cannot be edited by the user:** One such recent initiative comes from a Stanford University professor and Elsevier, which annually publishes a ranking that identifies the top 2% of the most influential researchers using data from Scopus.
- **Detection of fake GS profiles:** The authors of [23] present a machine learning-based method to detect misconfigured author profiles. GS can use these tools to detect and retract fake profiles.

**Author Contributions:** Conceptualization, L.M. and D.K.; Methodology, L.M. and D.K.; Investigation, L.M. and D.K.; Resources, L.M. and D.K.; Writing—Original Draft Preparation, L.M. and D.K.; Writing—Review & Editing, L.M. and D.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Hammarfelt, B. Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics and history. *Aslib J. Inf. Manag.* **2017**, *69*, 607–623. [\[CrossRef\]](#)
2. Ibrahim, H.; Liu, F.; Zaki, Y.; Rahwan, T. Google Scholar is manipulatable. *arXiv* **2024**, arXiv:2402.04607.
3. Katsaros, D.; Manolopoulos, Y. Impact and productivity of PhD graduates of computer science/engineering departments of Hellenic universities. *arXiv* **2017**, arXiv:1707.05801.
4. Prancutė, R. Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications* **2021**, *9*, 12. [\[CrossRef\]](#)
5. Prins, A.A.; Costas, R.; van Leeuwen, T.N.; Wouters, P.F. Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Res. Eval.* **2016**, *25*, 264–270. [\[CrossRef\]](#)
6. Moed, H.F.; Bar-Ilan, J.; Halevi, G. A new methodology for comparing Google Scholar and Scopus. *J. Inf.* **2016**, *10*, 533–551. [\[CrossRef\]](#)
7. Delgado López-Cózar, E.; Robinson-García, N.; Torres-Salinas, D. The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 446–454. [\[CrossRef\]](#)
8. Hvistendahl, M. China's publication bazaar. *Science* **2013**, *342*, 1035–1039. [\[CrossRef\]](#)
9. Burton, S.; Basil, D.Z.; Soboleva, A.; Nesbit, P. Cite me! Perspectives on coercive citation in reviewing. *J. Serv. Mark.* **2024**, *38*, 809–815. [\[CrossRef\]](#)
10. He, L.; Hausman, H.; Pajonk, F. Generative Artificial Intelligence: A New Frontier of Scientific Misconduct? *Int. J. Radiat. Oncol. Biol. Phys.* **2024**. [\[CrossRef\]](#)
11. Shaffer, E. Too many authors spoil the credit. *Can. J. Gastroenterol. Hepatol.* **2014**, *28*, 605. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Macdonald, S. The gaming of citation and authorship in academic journals: A warning from medicine. *Soc. Sci. Inf.* **2022**, *61*, 457–480. [\[CrossRef\]](#)
13. Oravec, J.A. The manipulation of scholarly rating and measurement systems: Constructing excellence in an era of academic stardom. *Teach. High. Educ.* **2017**, *22*, 423–436. [\[CrossRef\]](#)
14. Katsaros, D.; Akritidis, L.; Bozanis, P. The *f* index: Quantifying the impact of coterminal citations on scientists' ranking. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 1051–1056. [\[CrossRef\]](#)
15. Halevi, G.; Moed, H.; Bar-Ilan, J. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *J. Inf.* **2017**, *11*, 823–834. [\[CrossRef\]](#)
16. López-Cózar, E.D.; Robinson-García, N.; Torres-Salinas, D. Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. *arXiv* **2012**, arXiv:1212.0638.
17. Van Bevern, R.; Komusiewicz, C.; Niedermeier, R.; Sorge, M.; Walsh, T. *h*-index manipulation by merging articles: Models, theory, and experiments. *Artif. Intell.* **2016**, *240*, 19–35. [\[CrossRef\]](#)
18. Pavlou, C.; Elkind, E. Manipulating citation indices in a social context. In Proceedings of the International Conference on Autonomous Agents & Multiagent Systems (AAMAS), Singapore, 9–13 May 2016; pp. 32–40.
19. Van Bevern, R.; Komusiewicz, C.; Molter, H.; Niedermeier, R.; Sorge, M.; Walsh, T. *h*-index manipulation by undoing merges. *Quant. Sci. Stud.* **2020**, *1*, 1529–1552. [\[CrossRef\]](#)
20. Chawla, D.S. The citation black market: Schemes selling fake references alarm scientists. *Nature* **2024**, *632*, 966. [\[CrossRef\]](#)
21. Else, H. By Chatgpt Fool Scientists. *Nature* **2023**, *613*, 423. [\[CrossRef\]](#)
22. Hosseini, M.; Resnik, D.B. Guidance needed for using artificial intelligence to screen journal submissions for misconduct. *Res. Ethics* **2024**. [\[CrossRef\]](#)
23. Tang, J.; Chen, Y.; She, G.; Xu, Y.; Sha, K.; Wang, X.; Wang, Y.; Zhang, Z.; Hui, P. Identifying mis-configured author profiles on Google Scholar using deep learning. *Appl. Sci.* **2021**, *11*, 6912. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.