



HAL
open science

Assessing the Impact of Retracted Articles on the Reliability of Citing Literature: A Mixed-Based Proposal

Yagmur Ozturk, Frédérique Bordignon, Cyril Labbé, François Portet

► To cite this version:

Yagmur Ozturk, Frédérique Bordignon, Cyril Labbé, François Portet. Assessing the Impact of Retracted Articles on the Reliability of Citing Literature: A Mixed-Based Proposal. MET STI 2024, Nov 2024, On-line, France. 10.5281/zenodo.14142700 . hal-04791200

HAL Id: hal-04791200

<https://hal.science/hal-04791200v1>

Submitted on 19 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assessing the Impact of Retracted Articles on the Reliability of Citing Literature: A Mixed-Based Proposal

Yagmur Ozturk¹, Frédérique Bordignon^{2,3}, Cyril Labbé¹, François Portet¹

¹ Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

² Ecole des Ponts, 77455 Marne-la-Vallée, France

³ LISIS, INRAE, Univ Gustave Eiffel, CNRS, 77455 Marne-la-Vallée, France

Correspondence: yagmur.ozturk@univ-grenoble-alpes.fr

ABSTRACT

In this work, we explain our proposal to assess the impact of retracted papers on the reliability of articles that cite retracted literature. We plan to combine scientometrics and natural language processing (NLP) to estimate reliability. The ultimate goal is to create an automated framework to evaluate and flag potential reliability issues in scientific publications that cite retracted publications. As a first step, we collected and made available a dataset of metadata of articles that cite retracted literature.

Keywords: student submission, ongoing work, scientometrics, retractions, natural language processing, citations to retracted articles, scientific literature correction, scientific argument mining

INTRODUCTION AND BACKGROUND

Retraction is a process used to correct the scientific literature and to inform readers about publications that cannot be relied upon because they contain errors or have other issues (Wager et al., 2009). But retracted articles continue being cited as credible scientific sources and this can jeopardise scientific progress through propagation of erroneous information (Zhu et al., 2024). Citations to retracted articles (CRA) has been a widely researched subject in scientometrics, either in micro (e.g. analysing citations to highly cited retracted articles (Heibi & Peroni, 2021)) or macro studies (e.g. wide scale analysis of a certain scientific field to observe citations over time (Heibi & Peroni, 2022), (Dinh et al., 2019)). Although retractions are rare, the number of retracted articles has increased in recent years: Van Noorden (2023) reports that in 2023, more than 10k scientific publications were retracted, almost doubling the retractions recorded in 2022. To deal with the slow nature of retractions and to increase awareness, certain tools and platforms, such as Retraction Watch and the Problematic Paper Screener (PPS) (Cabanac et al., 2022) were built to, respectively, index and flag retracted and/or problematic publications. PPS's detector Feet of Clay (FoC) was specifically built to flag publications that cite retracted literature. It currently flags more than 720k publications citing at least one retracted article. Identifying and maintaining an updated record of these instances is crucial for raising awareness, which is a key function of the PPS. However, FoC flags problematic articles only by the presence of cited retracted articles, whether or not those articles are essential to the scientific contributions of the citing article. We would like to estimate the extent to which the reliability of a paper is affected by the fact that it cites retracted articles, be these papers retracted after publication or unfortunately cited without the (citing) author being aware of their retracted status. We aim to do this with a more nuanced approach recognising not all citations are equal, the context where they occur is important, and when it comes to retractions, the reason(s) why a paper has been retracted is/are of primary importance. To automatically obtain this estimation, we propose a combination of methods from scientometrics and NLP, which leverage different types of data, i.e., metadata, citation contexts (CC) and argument units (AU):

1. Metadata analysis provides valuable insights into the citing and cited articles such as publication dates, retraction reasons (RRs) and research fields (RFs). By leveraging metadata, we can identify several key factors that impact the reliability of citing literature, including: (i) RF congruence: we hypothesise that when citing and cited articles share the same RF, the citing literature may be more susceptible to reliability issues than inter-fields citations (ii) retraction density: the number of retracted articles cited within an article can influence its reliability, with higher densities potentially indicating greater concerns (iii) temporal relationships: the dates of retraction and publication have to be taken into account to distinguish between 'pre-retraction' citations (made before the retraction of the cited article) and 'post-retraction' citations (made after retraction). If the post-retraction citation context is used to criticise the retracted work, compare results with, or acknowledge the retraction, then this does not pose a problem for the citing article.
2. In NLP, we plan to use Scientific Argument Mining (SAM) methods to parse full-text of these articles, detect CC and classify sentences as AU (that may or may not contain citations) according to their argumentative function. SAM tasks can also include detection of relations between AUs (e.g. support or attack other AU). Large-scale application of SAM can be challenging due to data availability issues, such as accessing full-texts of articles that have indexed citations. Dealing with different citation conventions from various RFs and publishers (e.g. brackets, numbers, and superscripts) can also cause difficulties in text parsing. Setting aside such caveats, some baseline works that annotate articles for SAM tasks have shown large progress in detailed annotations of AUs (Teufel et al., 2009), coarse-grained annotations that merge detailed annotations (Lauscher et al., 2018b) and automated frameworks for SAM (Lauscher et al., 2018a, Binder et al., 2022).

This work-in-progress submission sets the stage for a new system to assess the severity of problems caused by CRA in citing articles, and to reassess their reliability. In the next sections, we outline the methodology for gathering and selecting metadata from articles likely to be affected by CRA. This foundational work paves the way for our future research, where we will leverage advanced NLP techniques to analyse the collected dataset. Ultimately, our goal is to create a robust framework for identifying and characterising the impact of CRA on citing articles, enabling more informed evaluation and reuse of scientific research.

METADATA COLLECTION FROM ARTICLES THAT ARE LIKELY TO BE AFFECTED BY CRA

To address the limitations in existing datasets, we compiled our own dataset of metadata from open-access articles that cite retracted articles. This collected dataset is designed to mitigate the caveats identified in the NLP literature discussed in the previous section, mainly targeting the gaps in metadata availability. We gathered the original data from FoC since it provides easy access to the DOIs of articles citing retracted articles. We applied a set of criteria while selecting articles: this filtering process aimed to identify cases where the content of an article is likely to be affected by a CRA and where we can successfully apply NLP tasks (e.g. accessing and parsing full-texts). Therefore, we apply a filter to obtain: (i) non-retracted citing articles (at the time of data curation) because we consider that corrections for retracted articles have already occurred, (ii) open access citing articles for full-text availability, (iii) cited retracted articles with at least one scientific content related reason of retraction, (iv) articles only (discrepancies between document types can be difficult for NLP tasks). The most important criterion is RRs since it gives direct information about how the cited retracted article can be affecting the citing article. RRs are curated by the Retraction Watch Database, and there are currently 108 reasons. We defined 29 of these as “scientific content related” reasons (e.g. unreliable results, data manipulation). We applied this criterion since we want to identify uncovered issues in the content of citing articles. If an article is retracted for other reasons that do not co-occur with scientific content related reasons, we assume that citation does not pose a problem for the citing article. In this final dataset, we collected the metadata of 4979 citing articles and 3193 cited retracted articles which resulted in 7424 citations. We gathered the following metadata in this dataset for further use in our estimation task: publishing dates, retraction date, RFs, RRs, DOIs. The published dataset can be accessed online (Ozturk, 2024).

COMBINING NLP METHODS WITH METADATA

We aim to detect CC within their AUs and classify them according to their functions. After this detection, we should then be able to identify the CC of retracted articles through DOIs. We aim for this pipeline to be easily applicable to full-texts of any scientific article, regardless of the RF. Depending on this classification, and the addition of collected metadata, we will provide a reliability estimation. For instance, the tool will raise the user's attention to the CC that needs to be checked because the citation was made after the retraction. Another example is when the citation does not have a very significant function, such as “background work” but was made multiple times which can signal a problem. A countercase would be if the CC has the function of criticism, which does not decrease the reliability of the citing article. We aim to test and present first results of our estimation task using two existing pipelines from Lauscher et al. (2018a) and Binder et al. (2022). The first pipeline offers different annotation layers, namely: argument identification (background claim, own claim, data), discourse role classification (Background, Unspecified, Challenge, Future Work, Approach, Outcome) and subjective aspect classification (None, Limitation, Advantage, Disadvantage-Advantage, Common Practice, Novelty). Since this system classifies every sentence of a given article, we aim to leverage information extracted from all of these annotation layers. Binder's pipeline was built on the argument identification scheme offered in Lauscher's, and showed a 7% F-score increase in argument detection and classification tasks. This is mainly due to their addition of SciBERT (Beltagy et al., 2019) token embeddings to the architecture. We expect that the three possible argument tags will not be enough to estimate impact as we defined in our task, but we plan to combine both of these annotations. Finally, with the addition of the metadata we will present first estimation results. We then plan to explore the caveats of these systems and provide more training data for the automatization tasks, such as annotation of articles from different RFs and accounting for imbalance in lesser used argument types.

CONCLUSION

We provided an overview, a dataset and explained certain caveats of a future assessment framework to detect the impact of retracted papers on the reliability of citing literature. We proposed a novel method based on both scientometrics and NLP. Assessing the severity of CRAs requires a nuanced approach which is difficult for the extent of the cases of CRAs in the scientific literature. We aim for this methodology to be largely applicable and contribute to human evaluation when necessary, including through post publication peer review when the estimated impact on reliability is high. With the dataset we built of eligible cases' metadata, we are currently working on the NLP tasks where we parse the full-text of these articles and classify the CC and arguments surrounding these citations. After this, we will again use the metadata to estimate the severity of problems, depending on the type of argumentation used, how many times retracted articles are cited, retraction and publication timelines and RFs.

ACKNOWLEDGEMENTS

We acknowledge the NanoBubbles project that has received Synergy grant funding from the European Research Council (ERC), within the European Union's Horizon 2020 program, grant agreement no. 951393 (<https://doi.org/10.3030/951393>).

REFERENCES

- Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 3615–3620). Hong Kong, China: Association for Computational Linguistics, DOI: 10.18653/v1/D19-1371
- Binder, A., Hennig, L., & Verma, B. (2022). Full-text argumentation mining on scientific publications. In T. Ghosal, S. Blanco-Cuaresma, A. Accomazzi, R. M. Patton, F. Grezes, & T. Allen (Eds.), *Proceedings of the first workshop on information extraction from scientific publications* (pp. 54–66). Online: Association for Computational Linguistics.
- Cabanac, G., Labbé, C., & Magazinov, A. (2022). The ‘Problematic Paper Screener’ automatically selects suspect publications for post-publication (re)assessment. *arXiv [Cs.DL]*. Retrieved from <http://arxiv.org/abs/2210.04895>
- Dinh, L., Sarol, J., Cheng, Y., Hsiao, T., Parulian, N., & Schneider, J. (2019). Systematic examination of pre- and post-retraction citations. *Proceedings of the Association for Information Science and Technology*, 56(1), 390–394. DOI: 10.1002/pr2.35
- Heibi, I., & Peroni, S. (2021). A qualitative and quantitative analysis of open citations to retracted articles: the wakefield 1998 et al.’s case. *Scientometrics*, 126(10), 8433–8470. DOI: 10.1007/s11192-021-04097-5
- Heibi, I., & Peroni, S. (2022). A protocol to gather, characterize and analyze incoming citations of retracted articles. *PLOS ONE*, 17(7), e0270872. DOI: 10.1371/Journal.pone.0270872
- Lauscher, A., Glavaš, G., & Eckert, K. (2018a). ArguminSci: A tool for analyzing argumentation and rhetorical aspects in scientific writing. In N. Slonim & R. Aharonov (Eds.), *Proceedings of the 5th workshop on argument mining* (pp. 22–28). Brussels, Belgium: Association for Computational Linguistics. DOI: 10.18653/v1/W18-5203
- Lauscher, A., Glavaš, G., & Ponzetto, S. P. (2018b). An argument-annotated corpus of scientific publications. In N. Slonim & R. Aharonov (Eds.), *Proceedings of the 5th workshop on argument mining* (pp. 40–46). Brussels, Belgium: Association for Computational Linguistics. DOI: 10.18653/v1/W18-5206
- Ozturk, Y. (2024). A Dataset of Metadata of Articles Citing Retracted Articles [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13621503>
- Teufel, S., Siddharthan, A., & Batchelor, C. (2009). Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In P. Koehn & R. Mihalcea (Eds.), *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 1493–1502). Singapore: Association for Computational Linguistics.
- Van Noorden, R. (2023). More than 10, 000 research papers were retracted in 2023 — a new record. *Nature*, 624(7992), 479–481. Retrieved from <http://dx.doi.org/10.1038/d41586-023-03974-8>, DOI: 10.1038/d41586-023-03974-8
- Wager, E., Barbour, V., Yentis, S., & Kleinert, S. (2009). Retractions: Guidance from the committee on publication ethics (COPE). *Croatian Medical Journal*, 50(6), 532–535. DOI: 10.3325/cmj.2009.50.532
- Zhu, H., Jia, Y., & Leung, S.-w. (2024). Citations of microrna biomarker articles that were retracted: A systematic review. *JAMA Network Open*, 7(3), e243173. DOI: 10.1001/jamanetworkopen.2024.3173