



HAL
open science

Bad smells in reviewers' reports? Text-mining the MDPI Open Peer Review Corpus

Gilles Hubert, Guillaume Cabanac, Cyril Labbé

► **To cite this version:**

Gilles Hubert, Guillaume Cabanac, Cyril Labbé. Bad smells in reviewers' reports? Text-mining the MDPI Open Peer Review Corpus. 2023. hal-04311568v1

HAL Id: hal-04311568

<https://hal.science/hal-04311568v1>

Preprint submitted on 28 Nov 2023 (v1), last revised 15 Nov 2024 (v2)

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.



Distributed under a Creative Commons Attribution 4.0 International License

Bad smells in reviewers' reports? Text-mining the MDPI Open Peer Review Corpus

[Gilles HUBERT](#),¹ [Guillaume CABANAC](#)^{1, 2} and [Cyril LABBE](#)³

¹ Université Toulouse 3 – Paul Sabatier, IRIT UMR 5505 CNRS, Toulouse, France

² Institut universitaire de France (IUF), Paris, France

³ Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Objective. Malpractice affecting the reviewing process is detrimental to science. We introduce methods to reveal evidence of peer review manipulation, such as template usage, citation manipulations or botched and meaningless reviewer reports. We apply and evaluate these methods on a corpus of reports.

Method. We downloaded the “MDPI Open Peer Review Corpus 2” webscraped by Miłkowski et al. (2023, <https://doi.org/10.18150/shkp7b>). We focused on ‘Round 1’ reports, retaining those in plain text (excluding those uploaded as attached files). We computed statistics on report length, identified references suggested by reviewers with regular expressions, extracted frequent word sequences, and analysed the pairs of reports showing an inter-textual similarity higher than 90%.

Results. The MDPI corpus consists of reports for 135,653 accepted articles in MDPI journals from 2011 to 2022. We mined this 170 GB dataset and observed several shortcomings. Some reports appear to be truncated due to failed webscraping. The dataset we analysed contains 135,437 articles and their 339,387 associated reports. The average report has 270 words with a median length of 202. Microscopic reviews consist of one word only, such as ‘accept’, ‘none’, ‘Nil’ or ‘N.A.’ ($n = 230$). These seem to be reports (mis)presented by the publisher as ‘Round 1’ albeit resulting from a peer review ran for an undocumented earlier submission (‘reject and resubmit’ editorial act). Tiny reports of less than 20 words account for 3.5% of the dataset. We also searched for report templates being constantly reused. Report–report similarities show that 40 reports were almost identical. At least 10 articles share two identical reports. Large chunks of text were reused across unrelated 380 reports sharing at least 300 words in chunks of 10+ words. We also spotted potential coercive citations: some reports contain the same wording suggesting authors to add a set of DOIs/PMIDs.

Conclusion. We found very few evidence of questionable features in the MDPI reports. More research is needed to improve malpractice detection and assess its prevalence in (open) peer review reports. Once implemented into publishers’ workflows, early misconduct detection should help to prevent botched reports, template reuse, and coercive citations.

Acknowledgements. GC and CL 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](#).

Objective: 50 words

Malpractices in the reviewing process are detrimental to the scientific literature. We present methods intended to find evidence of peer review manipulations, such as: template usage, citation manipulations or too short and meaning-less reviews. We test these methods on an open peer review corpus to assess its quality.

—> 48 mots

The scientific literature suffers from a variety of research integrity issues. Malpractices in scientific publications have been identified. Malpractices in the reviewing process would also be detrimental to the scientific literature.

We present methods intended to analyze peer reviews searching for evidence of manipulations, such as template usage, citation manipulations, or too short review.

-> 54 mots

Build and validate methods to find evidence of peer review manipulations: template usage (similarity), citation manipulations (DOI/PMID), too short review (none, NA, ...).

Method: 50 words

We downloaded the “MDPI Open Peer Review Corpus 2” (Miłkowski et al., 2023). We excluded reviews consisting solely of an attached file, considering only the ‘Round 1’ of the review process. We compute statistics on review lengths, search for suggested references using regular expressions, compute similarity scores and frequent n-grams.

-> 50 words

We applied the following methods on the plain text of reviews (i.e., excluding reviews consisting solely of an attached file) only for ‘Round 1’ of the review process, which are usually more extensive reports and thus more interesting to analyze:

- Compute statistics on review lengths;
- Search for suggested references based on regular expressions based on publication identifiers (i.e., PMID and DOI);
- Compute similarity scores between pairs of reviews;
- Search frequent n-gram (n=??) and construct graphs of reviews

-> 108 mots :-{

We downloaded the “MDPI Open Peer Review Corpus 2” (Miłkowski et al., 2023) covering 135,653 accepted articles (170 GB). Selected only “round 1” reports and excluded attached files.

- stat on length
- regexp PMID/DOI
- presence of tortured phrases
- compute similarity between round 1 reports
- compute frequent n-gram (n=??) and graph of review

Results: 200 words

The MDPI corpus consists of reviews for 135,653 accepted articles (170 GB) in MDPI journals from 2011 to 2022. The corpus has some inherent shortcomings. It contains incompletely recovered reviews and files attached to reviews are absent. For these reasons 13,670 reviews are excluded in our final corpus consisting of 339,387 reports about 135,437 distinct articles.

Some artefacts appear in the form of very small reviews consisting of only one word, such as 'accept', 'none', 'Nil' or 'N.A.'. They often correspond to reports (mis)presented by the publisher as 'Round 1' but are in fact the result of an additional round of an undocumented previous submission. We noticed 11751 such reviews (3.4%) smaller than 20 words. The 122,607 articles that have all reports in plain text have on average 270 words (median 202).

Nevertheless, the corpus provides multiple information. Although anecdotal with regards to the number of reviews, our study reveals questionable practices: at least 40 reviews entirely (or quasi-) identical, a minimum of 10 articles having two identical reviews, large chunks of text reused across unrelated reviews (380 reviews —1.1%— sharing at least 300 words in chunks of at least 20 words) and finally we are able to identify potential citation abuses.

—>197

, keeping entirely or partially the same reviewers, but without author responses to previous comments. We also found identical reviews for the same paper or for different papers, which question the review process.

Pb corpus (<https://pubpeer.com/publications/25353AAFD4FC52E2BEC8C7AD08B259>):

- web scraping issues reported to corpus makers (V1->V2)
- incomplete reviews (renamed attached files can't be matched to the reviews)
- artefact: small reviews (like "accept", "none"...) from assessment of resubmission that are miss presented by the publisher as "round1", Same review copy/paste several times (19x,...)

Features:

- 339,387 reports (+13,670 reviews éliminées: "see attached files"/pb scrapping) for 135,653 distinct articles
- 12,830 distinct articles with *fullreviews_False
- number of papers that have all reports in plain text: 122,607/135+
- stats: average of 225 (Med 165) words per review
- 11751 smaller than 20 words ("None", "Regards", "Best", "N.A.", "Nil"...)

Anecdotal evidence of questionable practices:

- Report Template or chunk of text are reused from one review to an other
- "Regenerate Response" ?? Not for now
- Citation abuse : Frequent PMIDs/DOIs, DOI/PMID suggestions for the journal where the paper is submitted
- Identical reviews for the same paper (probably an error in the editorial process?) or for different papers (manipulation of the review process)

Conclusion: 50 words

Thanks to our methods, we can identify questionable practices by reviewers and publishers. We found very few evident malpractices in reviewers' reports. Due to frequent false positive automatizing completely malpractice detection appears difficult. Quantifying precisely the phenomenon requires methods improvements.

-> 40

Thanks to our methods, we are able to identify questionable practices of reviewers and publishers. Very few evident malpractices in reviewers' reports appear. Due to false positive it is difficult to automatize completely the detection and to quantity precisely such phenomenon. Nevertheless there is a need for furthered methods.

—>49

Very few bad smells and few questionable practices through some anecdotal evidence. thanks to our methods we are able to identify questionable practices. Due to false positive it is difficult to automatize completely the detection and to quantity precisely the phenomenon. Nevertheless it shows that ... better tools are needed ... and publisher could already address some problems

—>58

References

Miłkowski, Marcin; Jasieński, Ksawery; Depta, Remigiusz, 2023, "MDPI Open Peer Review Corpus 2", <https://doi.org/10.18150/SHKP7B>, RepOD, V2

Bad smells in Reviewers' Reports: text-mining the MDPI Open Peer Review Corpus

Data

<https://doi.org/10.18150/SHKP7B>

Method

- computation of pairwise text similarity, visual inspection of highest similarity scores
- [GC] regexp to identify the most frequent DOIs and PMIDs

Results

- 'Regenerate Response' from ChatGPT
<https://pubpeer.com/publications/E6F750F5DE06F5C90B0455E1AB4563>
- Report = None, acknowledged by authors
<https://pubpeer.com/publications/1151A6FEE0ED92F4A02469D888AB4A>
- Empty report <https://pubpeer.com/publications/D8D8B9B3AEF07D3BE02593A2542A64>
- Identical reviews in different journals (review template)

https://www.mdpi.com/2673-7116/2/2/17/review_report Reviewer 2

https://www.mdpi.com/2071-1050/14/7/4068/review_report Reviewer 1

https://www.mdpi.com/2076-3417/11/23/11187/review_report Reviewer 1

https://www.mdpi.com/2071-1050/14/2/1016/review_report Reviewer 2 (Anonymous)

https://www.mdpi.com/2076-3417/12/19/9814/review_report Reviewer 2 (Ali A. Shubbar)

https://www.mdpi.com/2076-3417/12/17/8508/review_report Reviewer 2 (Ali Shubbar)

https://www.mdpi.com/2073-445X/11/2/312/review_report Reviewer 3

https://www.mdpi.com/2071-1050/13/3/1288/review_report Reviewer 2

the following differing only from suggested references:

biomedicines10020240 https://www.mdpi.com/2227-9059/10/2/240/review_report

biomedicines10061340 https://www.mdpi.com/2227-9059/10/6/1340/review_report

cancers13215505 https://www.mdpi.com/2072-6694/13/21/5505/review_report

cancers14010072 https://www.mdpi.com/2072-6694/14/1/72/review_report

cancers14040865 https://www.mdpi.com/2072-6694/14/4/865/review_report

cancers14071680 https://www.mdpi.com/2072-6694/14/7/1680/review_report

cancers14071730 https://www.mdpi.com/2072-6694/14/7/1730/review_report

cancers14102386 https://www.mdpi.com/2072-6694/14/10/2386/review_report

cancers14122838 https://www.mdpi.com/2072-6694/14/12/2838/review_report

curroncol29050266 https://www.mdpi.com/1718-7729/29/5/266/review_report

curroncol29060346 https://www.mdpi.com/1718-7729/29/6/346/review_report

ijms23010223 https://www.mdpi.com/1422-0067/23/1/223/review_report

ijms23052704 https://www.mdpi.com/1422-0067/23/5/2704/review_report

ijms23052801 https://www.mdpi.com/1422-0067/23/5/2801/review_report

ijms23158133 https://www.mdpi.com/1422-0067/23/15/8133/review_report

jcm10081638 https://www.mdpi.com/2077-0383/10/8/1638/review_report
jcm11051213 https://www.mdpi.com/2077-0383/11/5/1213/review_report
jpm12040540 https://www.mdpi.com/2075-4426/12/4/540/review_report

- Some DOIs/PMIDs are (too) frequently suggested (citation abuse?)
- [Cyril] Frequent ngrams — histogramme des tailles — détection des tortured phrases
- Reports des papiers rétractés ? <https://www.mdpi.com/2076-3417/9/11/2316> (trouvé https://www.mdpi.com/2076-3417/9/11/2316/review_report)
- jointure avec torturedpapers: uniquement un cas <https://doi.org/10.3390/jcp1020017>

—

From other people:

- <https://twitter.com/clementFFF/status/1706987985764270231> Final decision Reject

— Accept : 26

— Accepted: 7

— None : 59

— No comments : 70

— Nb de reveiw < 20 mots 11751