

A qualitative and quantitative citation analysis toward retracted articles: a case of study

Ivan Heibi – ivan.heibi2@unibo.it – <https://orcid.org/0000-0001-5366-5194>

Research Centre for Open Scholarly Metadata, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Digital Humanities Advanced Research Centre (/DH.arc), Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Silvio Peroni – silvio.peroni@unibo.it – <https://orcid.org/0000-0003-0530-4305>

Research Centre for Open Scholarly Metadata, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Digital Humanities Advanced Research Centre (/DH.arc), Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

Abstract

In this article, we show the results of a quantitative and qualitative citation analysis on a popular and highly cited retracted paper: “Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children” by Wakefield et al., published in 1998. The main purpose of our study is to understand the behavior of the publications citing retracted articles and the characteristics of the citations the retracted articles accumulated over time. Our analysis is based on a methodology which illustrates how we gathered the data, extracted the topics of the citing articles, and visualised the results. Such results concerned the analysis of the entities citing Wakefield et al. and their related in-text citations. We observed a constant increasing number of citations in the last 20 years, accompanied with a constant increment in the percentage of those acknowledging its retraction. Citing articles have started discussing/dealing with the retraction of Wakefield et al. even before its full retraction, happened in 2010. Social sciences studies were among those that have mostly discussed the retraction phenomena. In addition, when observing the in-text citations, we noticed how a large part of the citations received by Wakefield et al.’s article have been focused on general discussions without recalling strictly medical details, especially after the full retraction. Medical studies did not hesitate in acknowledging the retraction and even gave strong negative statements on it in some occasions.

Keywords: citation analysis, retraction, topic modeling, Science of Science

Introduction

A peer-reviewed retracted article should be considered as an invalid source of knowledge depending on specific reasons which might include scientific misconduct, fabrication, general content errors, plagiarism and self-plagiarism (Moylan & Kowalcuk, 2016). A retraction is established by the editorial board of the venue in which the original publication was published, and it is accompanied by a retraction notice. Also, sometimes a meaningful label (e.g. "RETRACTED") is associated with the retracted article either in the article title or as a label specified upon the article content.

In order to make retractions more visible to a reader, existing services, such as CrossMark by Crossref, have been proposed in the past years to show updated notifications regarding articles, such as retractions and error corrections (Meyer, 2011). An important service which keeps track and collects retractions of scholarly articles is Retraction Watch (<http://retractionwatch.com/>) (Collier, 2011). However, usually, retracted articles are not deleted from the publisher's catalogue, to enable future research on them.

The retraction phenomenon has been largely discussed by scientometricians. We can organise the studies in this domain into two macro categories: (a) large scale analysis, (b) case of study analysis.

Works belonging to category (a) focus on either an analysis of a single field of study or a broader domain, such as a macro area. Usually, these studies try to answer general questions such as how retractions influence the impact on the authors, institutions and the retracted work itself.

Large scale citation analysis on retracted articles have been mostly focused on quantitative aspects. For instance, Lu et al. (2013) used the citation data collected from Web Of Science to demonstrate that a single retraction could trigger citation losses through an author's prior body of work. The negative repercussions on authors and co-authors of retracted articles have been demonstrated also by other works such as (Azoulay et al., 2017) (Mongeon & Larivière, 2016) (Shuai et al., 2017). Others have discussed possible approaches to avoid retraction and related issues in citing retracted papers. For example, Mott et al. (2019) suggest strategies to adopt for improving the effectiveness of retraction notices and a better conscious citing.

The studies of category (b) consider either single or multiple retracted article cases (usually, popular cases) and perform content analysis of the articles citing retracted ones. Generally, their main goal is building a general approach to apply on a larger scale starting from the findings and results obtained – e.g. by focusing on post-retraction citations and the related sentiment when citing (Bar-Ilan & Halevi, 2017), by classifying citation contexts (Jan et al., 2018), and by running a network analysis study (Chen & Leydesdorff, 2014). Similarly, the work done by Teixeira da Silva & Dobránszki (2017) focuses on a restricted list of the top ten cited retracted articles for analysing the number of citations and retraction reasons, without considering the content of the articles citing the retracted ones.

Other studies of category (b), instead, focus only on one specific retraction case. The aims of these works are to observe where in-text citations to the retracted article appear in the text of the citing articles (van der Vet & Nijveen, 2016), to notice that retracted works are still

being cited without mentioning their retraction (Bornemann-Cimenti et al., 2016), and to classify reasons for citing the retracted article (Luwel et al., 2019).

The analysis we present in this article is close to those introduced in the latter set of studies. In particular, we want to focus on a highly cited retracted article, i.e. (Wakefield et al., 1998), that suggested a link between autism and childhood vaccines. This article was partially retracted in 2004, and subsequently fully retracted in 2010. According to an analysis conducted by Retraction Watch, this retracted article was the third most cited article, based on a citation count retrieved in Web of Science (<https://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/>) as of May 2019. Throughout our article, we refer to it with the abbreviation WF-PUB-1998.

We think that the retraction of WF-PUB-1998 is an important case that deserves to be analysed considering its popularity among several anti-vaccines movements and the relevant implications it had (and, sometimes, still has) for society. It has been recently analysed also by Suelzer et al. (2019) who showed that the large part of the articles citing WF-PUB-1998 deny its conclusions, although a significant number of them did not document its retraction.

In our study, we focus on the citation analysis of the WF-PUB-1998 from a quantitative and qualitative point of view. We relied on open and free services to retrieve the articles (and their metadata) citing WF-PUB-1998, and we used automatic natural language processing techniques to conduct a qualitative study on the content of the article citing WF-PUB-1998. In particular, the aim of our work is to answer to the following research questions:

- RQ1) What are the research topics introduced in the citing articles citing WF-PUB-1998 before and after its retraction?
- RQ2) What are the most relevant characteristics of the in-text citations in the articles citing WF-PUB-1998 before and after its retraction?

Methodology

The methodology of this work is based on three different steps. The first two steps (subsection “Data gathering” and “Topic modeling”) define the methods for annotating and generating the data that we need for our study, while the third step (subsection “Addressing the research questions”) defines how we try to answer RQ1 and RQ2.

Data gathering

The data sources we used to gather the data for our analysis were OpenCitations COCI (Heibi et al., 2019) (<http://opencitations.net/index/coci>), that we used to retrieve citation data, the RetractionWatch database (<http://retractiondatabase.org/>) used to retrieve information of retracted articles, and SCImago (<https://www.scimagojr.com/>) to retrieve subject areas and subject categories of publications.

We queried the COCI REST API (<http://opencitations.net/index/coci/api/v1>) when COCI was populated with citation data of its November 2018 release (OpenCitations, 2018), that

contained 445,826,118 citation links coming from 46,534,705 bibliographic resources. Among the attributes that COCI uses for characterizing each citation having WF-PUB-2018 as cited entity, we took into consideration the citing DOI, the cited DOI, and the creation date of the citation (i.e. the publication date of the citing entity).

This stage was organised in five steps, introduced in Table 1. Additional information and details about the workflow of this step of our methodology can be found in (Heibi & Peroni, 2020).

Step	Description	Input	Output
1) Identifying and retrieving the citing entities	Identifying the list of entities citing WF-PUB-1998 and storing their main metadata	DOI of the retracted article	For each citing entity: 1.1) <i>DOI</i> 1.2) <i>year of publication</i> 1.3) <i>title</i> 1.4) <i>venue id (ISSN/ISBN)</i> 1.5) <i>venue title</i>
2) Retrieving the citing entities characteristics	Annotating whether the citing entities have been or have not been retracted as well	DOIs of the citing entities	For each citing entity: 2.1) <i>is / is not retracted</i>
3) Classifying the citing entities according to subject areas and subject categories	Classifying the citing entities into areas of study and specific subject categories, following the SCImago classification	ISSN/ISBN of publication venues of citing entities	For each citing entity: 3.1) <i>subject area</i> 3.2) <i>subject category</i>
4) Extracting textual values from the citing entities	Extracting the citing entities' abstracts, the in-text reference pointers, citation contexts, title of the section where the in-text citations happen	DOIs of the citing entities	For each citing entity: 4.1) <i>abstract</i> 4.2) <i>in-text citation section</i> 4.3) <i>in-text citation context</i> 4.4) <i>in-text reference pointer</i>
5) Annotating the in-text citations characteristics	Annotating the intent and sentiment of each in-text citation, and specifying whether the text in citation contexts mentions the retraction of the cited article	In-text citations' contexts	For each in-text citation: 5.1) <i>citation intent</i> 5.2) <i>citation sentiment</i> 5.3) <i>retraction is / is not mentioned</i>

Table 1. An overview of all the steps needed for generating an annotated dataset of WF-PUB-1998 citing entities, to be further used during this work. For each step, we provide a brief description, the inputs needed, and the output produced. The output is represented as the list of features that will be included in the final dataset used for our analysis.

Gathering raw data

We retrieved the DOI, year of publication, title, ISSN/ISBN of the publication venue and the related title of all the articles citing WF-PUB-1998 starting from its DOI. For doing that, we

used the “citations” operation of the OpenCitations COCI APIs (<http://opencitations.net/index/coci/api/v1#/citations/{doi}>) to get the list of all citing entities, then we used the “metadata” operation (<http://opencitations.net/index/coci/api/v1#/metadata/{dois}>) to get the metadata of each citing entity.

Then, we queried the RetractionWatch database (<http://retractiondatabase.org/>) to manually check if each of the citing entities (identified by its DOI) has been retracted itself or not, and we identified the subject areas and subject categories of each citing entity using the identifiers (either ISSN or ISBN) of the publication venue of the cited entity. For publication venues with ISSN, we used the SCImago Journal Rank service (<https://www.scimagojr.com/>). In case a venue had more than one subject area or subject category, we considered all of them. For publication venues with ISBN, we used the ISBNDB service (<https://isbndb.com/>) to look up the related Library of Congress Classification (LCC, <https://www.loc.gov/catdir/cpsu/lcco/>). Then, we mapped the LCC categories we found to SCImago subject areas as follows:

1. We considered only the starting alphabetic segment of the LCC code and find the corresponding LCC discipline using a pre-built lookup index (e.g. “RC360” -> “RC” -> “Medicine”).
2. We checked whether the value of the LCC subject matches the exact value of a Scimago area using a pre-built Scimago mapping index. If the corresponding value is present, the algorithm will automatically annotate the subject area with such value, while the subject category will have the same value with the addition of “(miscellaneous)” at the end of it. This is usually done on the Scimago classification to express a general category of a specific area of study. In case no corresponding Scimago area has been found, we continued to point 3.
3. We checked whether the value of the LCC subject is a Scimago category using the same pre-built Scimago mapping index. If the corresponding value is present, we annotated the corresponding category with such value, while the area will have the same value used on the Scimago classification to denote the macro area of such category. In case no corresponding Scimago category was found, we continued to point 4.
4. The remaining ISBN values need to be manually annotated through the consultation of the complete LCC index ([http://www.loc.gov/catdir/cpsu/lcco/](https://www.loc.gov/catdir/cpsu/lcco/)).

Finally, starting from the DOI of the citing entities, we retrieved the full-text of all the citing articles. From the full-text of such articles, we extracted their abstracts, the in-text reference pointer denoting a bibliographic reference referencing WF-PUB-1998 (e.g. “Wakefield et al., 1998”), the citation contexts of in-text citations and the sections where the citation contexts were contained.

For this study, we defined the in-text citation contexts as the sentence that contains the in-text reference pointer to WF-PUB-1998 (i.e. the anchor sentence), plus the preceding and following sentences. Special cases/exceptions to this rule are fully introduced in (Heibi & Peroni, 2020). Also, we characterised each of the sections containing in-text citations according to its type – using the categories “introduction”, “method”, “abstract”, “results”,

“conclusions”, “background”, and “discussion” listed in (Suppe, 1998). These categories have been used when the intent of the section was clear, otherwise we used other three residual categories, i.e. “*first section*”, “*final section*” and “*middle section*” combined with the original title of the section. If the examined full-text of the citing entity is not organized into sections/paragraphs, then the value of its in-text citation section is set to “*none*”. For instance, this could be the case for citing entities that are editorials.

Annotating the in-text citations

We analysed each citation context of the in-text citation retrieved, and we inferred:

- the author's sentiment regarding WF-PUB-1998;
- whether at least one citation context of any in-text citation of the citing entity does explicitly mention the fact that the cited entity has been retracted (i.e. the citation context contained the word “retract” or one of its derivative words – “retractions”, “retracted”, etc.);
- the citation intent (or citation function), defined as the authors' reason for citing a specific article (e.g. because the citing entity uses a method defined in the cited entity).

For specifying the author sentiment, we followed the classification proposed by (Bar-Ilan & Halevi, 2017). Thus, we annotated each in-text citation with one of the following values:

- positive, when the retracted article was cited as sharing valid conclusions, and its findings could have been also used in the citing study;
- negative if the citing study cited the retracted article and addressed its findings as inappropriate and/or invalid;
- neutral, when the author of the citing article referred to the retracted article without including any judgment or personal opinion regarding its validity.

To record the citation intent, we used the citation functions specified in the Citation Typing Ontology (CiTO, <http://purl.org/spar/cito>) (Peroni & Shotton, 2012), an ontology for the characterization of factual and rhetorical bibliographic citations. Even if, in principle, an in-text citation might refer to more than one CiTO function at the same time, we decided to annotate each in-text citation with one citation function only.

We performed the annotation of the in-text citations using the decision model we developed for this study, summarised in Figure 1. This decision model is based on a priority ranked strategy that works as follows:

1. we matched each in-text citation to WF-PUB-1998 against at least one of the three macro-categories, i.e. “Reviewing”, “Affecting” and “Referring” (first row in Figure 1);
2. for each macro-category selected, we selected one or more citation functions choosing between those provided by CiTO;
3. if we selected only one citation function, we annotated the in-text citation intent with such a function, otherwise
4. we calculated the priority of each citation function selected by summing its value in parenthesis (from 0.1 to 0.6) with the corresponding value defined in the y-axis (from 10 to 50) and in the x-axis (from 1 to 8), as shown in Figure 1. The smaller the sum,

the more priority the citation function has. For instance, the priority of the citation function “confirms” is 11.2 that is higher than the one of the citation function “describes”, which is 43.2. Finally, we selected the citation function that has higher priority and annotated the in-text citation function with it.

Reviewing and eventually giving an opinion on the cited entity			Affecting either the content or the perception toward the cited/citing entity		Referring to the cited entity for material/conceptual purposes			
10	Fill in the sentence: "My statements are <u>HEADER</u> , the cited entity, such that they <u>CITO-citation-function</u> "		Fill in the sentence: "My statements <u>CITO-citation-function</u> the cited entity, and affect the content of/perception toward the <u>HEADER</u> ."		Fill in the sentence: "The document I am citing represents a <u>HEADER</u> , such that my statements <u>CITO-citation-function</u> the cited entity"			
	E.g. "My statements are <u>Not on the same page with</u> the cited entity, such that they <u>critiques</u> "		E.g. "My statements <u>corrects</u> the cited entity, and affect the content of/perception toward the <u>Cited entity</u> ."		E.g. "The document I am citing represents a <u>General source</u> , such that my statements <u>cites for information</u> the cited entity"			
10	(0.1) supports (0.2) confirms	(0.1) derides (0.2) ridicules (0.3) refutes (0.4) critiques						
20	(0.1) agrees with	(0.1) disagrees with (0.2) disputes	(0.1) compiles (0.2) retracts (0.3) replies to (0.4) speculates on (0.5) corrects (0.6) extends	(0.1) uses data from (0.2) uses method in (0.3) uses conclusions from (0.4) obtains support from				
30		(0.1) parodies (0.2) qualifies (0.3) credits	(0.1) updates	(0.1) obtains background from				
40		(0.1) discusses (0.2) describes		(0.1) includes quotation from				
50				(0.1) includes excerpt from (0.2) documents (0.3) reviews	(0.1) cites as metadata document (0.2) cites as data source (0.3) cites as source document	(0.1) cites as authority (0.2) cites as evidence (0.3) cites as potential solution (0.4) cites as recommended reading (0.5) cites as related	(0.1) cites for information	
	1	2	3	4	5	6	7	8

Figure 1. The decision model for the selection of a CiTO citation function to use for the annotation of the citation intent of a an examined in-text citation based on its context

Topic modeling

We run a topic modeling analysis on the textual features gathered (i.e. abstracts and citation contexts). We use MITAO (Ferri et al., 2020) (<https://github.com/catarsi/mitao>), a visual interface to create a customizable visual workflow based on the Latent Dirichlet Allocation (LDA) topic modeling (Jelodar et al., 2019).

A standard workflow for building a topic model is composed of three main steps. Tokenization is the process of converting the text into a list of words, by removing punctuations, unnecessary characters, and stopwords. Stopwords also included, for abstracts, tokens used in structured abstracts such as “background”, “summary”, and “results”, and, for citation contexts, tokens used in the bibliographic reference of WF-PUB-1998 such as “Wakefield”, “Ileal”, and “lymphoid”. Since topic modeling can drastically benefit from the lemmatization (May et al., 2019), we decided to lemmatise all the tokens obtained by processing the abstracts and citation contexts.

Then, we created vectors for each of the tokens retrieved. In particular, we used the term frequency-inverse document frequency (TF-IDF) model to vectorize our words. The TF-IDF model takes into account the importance of the words based on its rarity in the document (i.e. either the abstract or the citation contexts). This model is considered as a good word weighting schema for general purpose textual collections and when the frequent terms may not be that much representative of the document topics (Bengfort et al., 2018) (Truica et al., 2016).

Finally, we built two topic models, one for the abstracts and one for the in-text citations context. In order to decide about the right number of topics to consider in each case, we computed and used the topic coherence score, as suggested in (Schmiedel et al., 2019). This score measures the degree of semantic similarity between high scoring words in the topic, and it helps us distinguish between topics that are semantically interpretable and topics that are artifacts of a mere statistical inference. Thus, for each of our cases (abstracts and citation contexts), we calculated the average coherence score for a range of models trained with a different number of topics (from 1 topic to 40 topics). Then, we plotted these values, we observed the number of topics for which the average score plateaued, and we selected a number of topics indicated in the plateau. For instance, Figure 2 shows the coherence score values of different LDA topic models built with a number of topics ranging from 1 to 40 using the citations contexts. The coherence score plateaued around 22-23 topics. Thus, we decided to look for 22 topics for the citation contexts. A similar approach has been also used for abstracts.

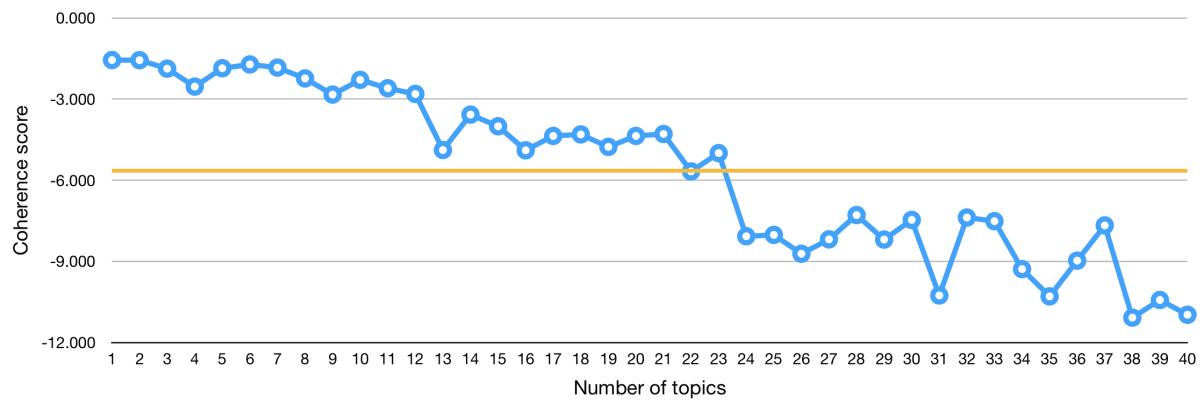


Figure 2. The coherence score of different LDA topic models built using a variable number of topics, from 1 to 40. The topic model is based on the corpus and dictionary of the in-text citations contexts. The orange line is the average value, and it plateaus around 22-23 topics.

After choosing the right number of topics, we used MITAO to generate two datasets for each case (abstracts and citation contexts):

- the 30 most important keywords of each topic, which represent the 30 most useful and probable terms for interpreting a topic, ranked according to their probability value;
- documents representativeness, i.e. the lists of all the documents of the corpus and their representativeness against each topic.

We also used MITAO for generating two interactive visualizations which we used to highlight important aspects of our study: LDavis and MTMvis.

LDavis provides a graphical overview of the topic modeling results (Sievert & Shirley, 2014). This visualization plots the topics as circles in a two-dimensional plane whose centers are determined by computing the distance between topics and uses a multidimensional scaling to project the inter-topic distances onto two dimensions. The topic prevalence is represented by the dimension of the area of each circle. LDavis shows a global list of 30 terms ranked using the “term saliency” measure. This saliency measure combines the overall probability of

a term with its distinctiveness: how informative is a specific term for determining the generation of a topic, versus any other randomly-selected term (Chuang et al., 2012). In addition, one can select a singular topic and LDAvis will show a list of 30 terms ranked using the “relevancy” measures. We used the default relevancy metric as defined in (Sievert & Shirley, 2014) to show the ranking of terms according to their topic-specific probability.

MTMvis (Metadata-based Topic Modeling Visualization) provides an interactive visualization which shows the representativeness of the topics in the documents based on a customisable metadata specified for those documents. We created two visualisations for both the abstracts and the citation contexts based, respectively, on the year of publication and the subject area of the articles citing WF-PUB-1998.

In Figure 3 we show the MITAO workflow we used for generating the above datasets and visualizations.

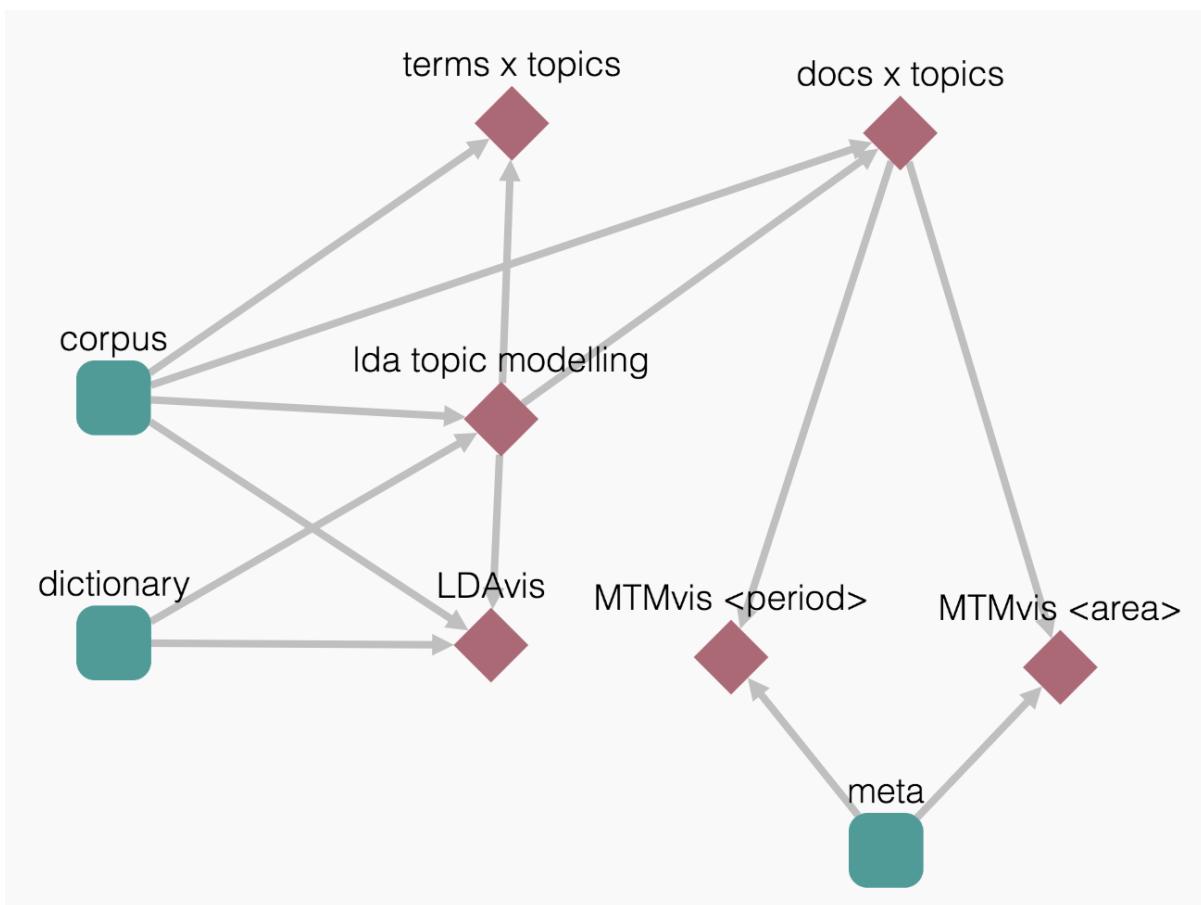


Figure 3. The MITAO workflow used for building a topic model and generating: (a) the datasets (“terms x topics” and “docs x topic”), and (b) the visualizations (“LDavis”, “MTMvis <period>”, and “MTMvis <area>”). The workflow takes three inputs: (a) the vectorized corpus (“corpus”), (b) a dictionary of words based on the tokenization results, and (c) the metadata of the original documents forming the corpus.

Addressing the research questions

To answer RQ1 and RQ2, we investigated the citing entities data and combined such data with the results of the topic modeling process.

Our approach takes into consideration the years of the partial (2004) and final (2010) retraction of WF-PUB-1998 to define three periods: (P1) from WF-PUB-1998 publication to the partial retraction (years: 1998-2004), (P2) from the partial retraction and to the final retraction (years: 2005-2010), and (P3) from the final retraction to 2017 (years: 2011-2017).

We used LDAvis and MTMvis to analyse the results of the topic model obtained. On the one hand, we used LDAvis to have a general overview of the topics, inspect their prevalence, and their terms. On the other hand, we used MTMvis to plot the corpus documents' topic representativeness.

Regarding RQ1, we mainly needed to analyse the results obtained by the topic model of the abstracts. The idea was to monitor the evolution of the emerging topics considering the three periods P1-P3 to show the main arising changes. We compared these observations against the area of study to highlight the evolution of citing behaviour in each individual area.

When dealing with RQ2, we primarily considered the features which characterized the in-text citations, such as the citation intent and the sentiment. The idea was to analyse these features against the outcomes of the topic model of the in-text citations contexts.

Results

In this section, we present the results of our analysis. All the data and visualizations are available in (Heibi, Ivan & Peroni, Silvio, 2020). Although in this article we present a screenshot of the visualizations, these are provided in dynamic HTML documents, and each visualization can be customised using the filters and parameters it makes available. We provided a dedicated webpage (<https://ivanhb.github.io/ret-analysis-wakefield-results/>) to enable readers to use such dynamic visualizations that we present in this work.

We organise the presentation of the results in two sections describing (a) the entities citing WF-PUB-1998 and (b) their in-text citations to WF-PUB-1998. For both, we introduce the data and the features we used for the analysis and then we present the outcomes of the related topic models.

Citing entities

The total number of citing entities gathered is 615. In Table 2, we list all the features we collected related to the citing entities. In particular, the first column lists the features with a brief description, while the second column summarizes its values, the total number of citing entities having such values and, if applicable, a classification of the different possible values.

WF-PUB-1998 citing entities features	Values
<i>doi</i> The DOI of the citing article	Total: All the citing entities had a value specified.
<i>year</i> The year of publication of the citing article	Total: All the citing entities had a value specified. Values: From 1998 (year of publication of WF-PUB-1998) to 2017.
<i>title</i> The title of the citing article	Total: All the citing entities had a value specified.

source_id The ID (ISSN/ISBN) of the venue of publication of the citing article	Total: 599 (97%) citing entities had a value specified. Values: ISSNs (548), ISBNs (51).
source_title The title of the venue of publication of the citing article	Total: 603 (98%) citing entities had a value specified.
retracted A yes/no value depending on whether the citing article has or has not received at least one retraction notification.	Total: 1 citing entity.
area The subject areas of the venue of publication of the citing article, based on the the SCImago Journal Classification (https://www.scimagojr.com/)	Total: 576 (93%) citing entities had at least a value specified. Values: 24 different values: "medicine" (380), "social sciences" (90), "nursing" (81), "biochemistry, genetics and molecular biology" (59), "psychology" (58), "pharmacology, toxicology and pharmaceuticals" (54), "immunology and microbiology" (52), "arts and humanities" (28), "neuroscience" (24), "environmental science" (17), "agricultural and biological sciences" (16), "health professions" (15), "computer science" (13), "mathematics" (10), "business, management and accounting" (8), "engineering" (7), "dentistry" (7), "multidisciplinary" (7), "decision sciences" (7), "economics, econometrics and finance" (5), "earth and planetary sciences" (1), "chemical engineering" (1), "materials science" (1), "physics and astronomy" (1)
category The subject categories of the venue of publication of the citing article, based on the the SCImago Journal Classification (https://www.scimagojr.com/)	Total: 576 (93%) citing entities had a value specified. Values: 170 different values.
abstract The abstract of the citing article	Total: All the citing entities had a value specified.
mention_retraction A yes/no value that indicates if at least one of the citation contexts of the citing article explicitly mentions the fact that the cited entity is retracted.	Total: All the citing entities had a value specified. Values: no (464), yes (151)

Table 2. The features that directly characterize the citing entities. The first column lists the features with a brief description, while the second column summarizes the related values we gathered.

Figure 4 introduces some descriptive statistics of the values described in Table 2. The charts are organised in three distinct rows, one for each period considered (P1-P3), mentioned in the first column. The second column contains the distribution per year of the citing articles according to the fact they either mention the retraction of WF-PUB-1998 (in green) or they do not (in red). On top of each bar in the chart, we also specify the number of citing entities the bar refers to. The third column contains the subject areas of the citing entities. The chart shows the ten most represented areas of study, while it groups all the other values (if any) in the last slice of the pie with the “Others” label.

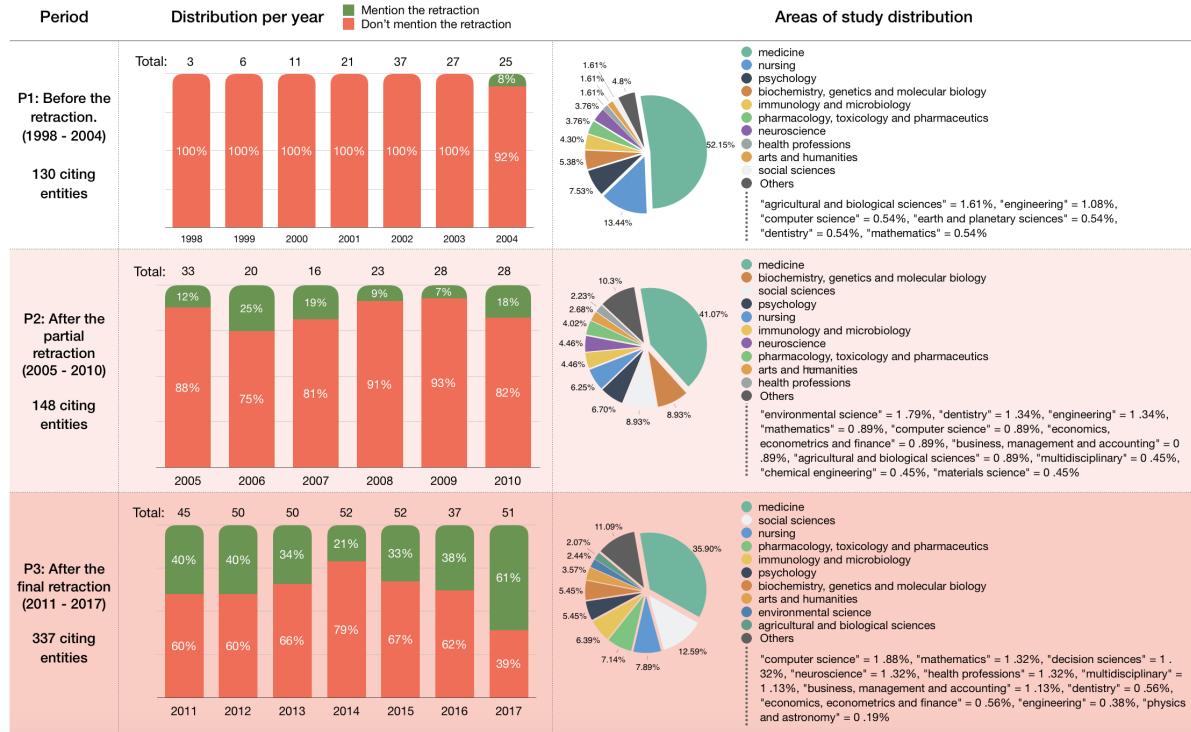


Figure 4. A summary of the citing entities. The first column contains the periods P1-P3 we considered, the second column shows the distribution per year of the citing entities that do mention (in green) or do not mention (in red) the retraction of WF-PUB-1998, while the third column shows the distribution of the subject areas of the citing entities.

The topic model we obtained from these entities is entirely based on their abstracts. Considering the results of the coherence score introduced in Section “Topic Modeling”, we built a topic model on 13 topics. Figure 5 shows the related LDAvis visualization. The left part of it shows two different clusters, and one of the clusters is composed of one big topic, i.e. topic-3, which was by far the larger topic identified by the process. Looking at the 30 most salient terms, the term “retract” is in the 5th position, meaning that some of the citing entities talked about the retraction of WF-PUB-1998 or, more generally, the retraction phenomena. The same list includes terms such as “social”, “movement”, “debat”, “media” and “cultur” which seem not to be strictly related with medical jargon. This scenario may be an indicator that some of the citing entities are not medical publications. Finally, among these 30 most salient terms, we found terms with a strong negative connotation, such as “fraud”.

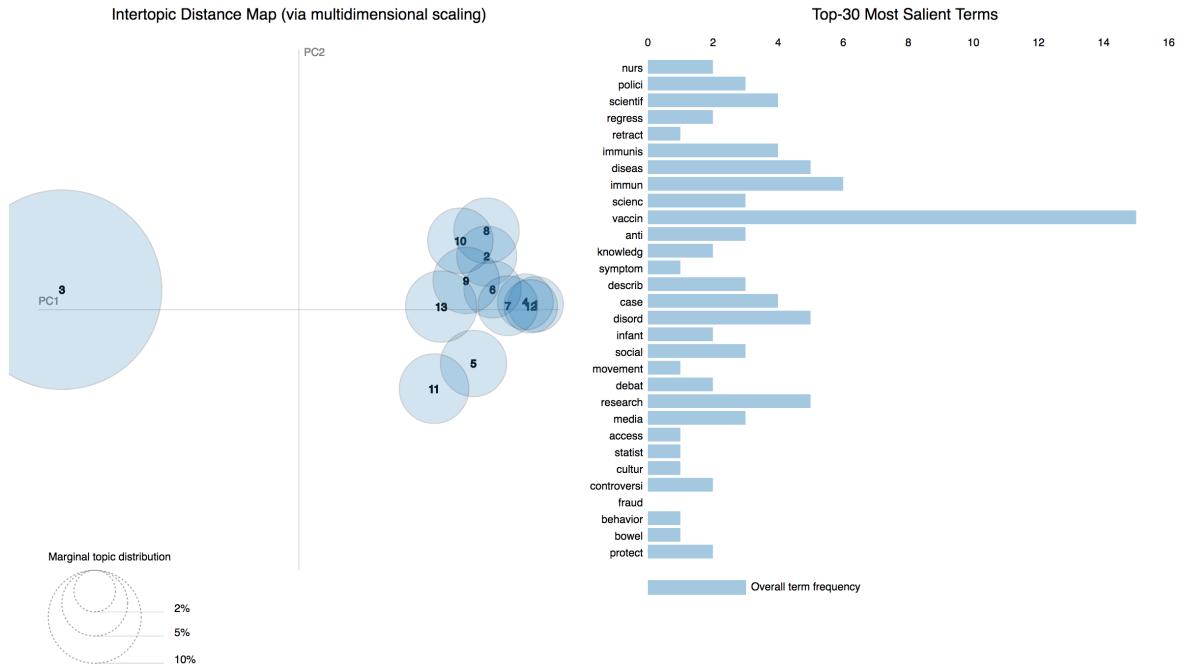


Figure 5. The LDAvis visualization built over the topic model obtained from the abstracts of the citing entities.

Using the data obtained through the topic model, we were able to explore each individual topic and give a possible interpretation to it by analysing its 30 most probable terms, as shown in Table 5 (in Appendix).

The MTMvis visualizations are plotted considering the period P1-P3 (Figure 6) and the subject areas of the citing articles (Figure 7). As shown in Figure 6, the topics 1, 2, and 5 were constantly increasing their percentages over the time while, on the contrary, topics 4 and 9 were decreasing. Along the same lines, topics 3 and 11 showed a very similar pattern along the three periods. As shown in Figure 7, some subject areas, such as *medicine* and *social sciences*, referred to almost all the topics while others (e.g. *computer science*) referred to particular subset of topics.

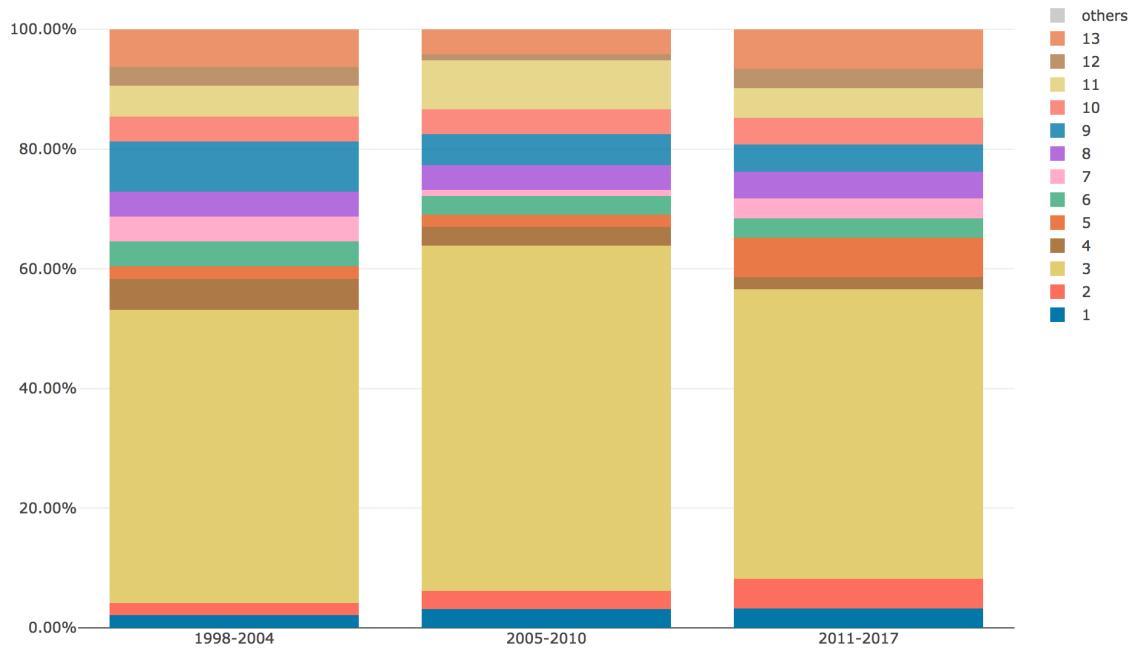


Figure 6. MTMvis built on the topic model obtained from the abstracts of the citing entities, shown against the three period P1-P3.

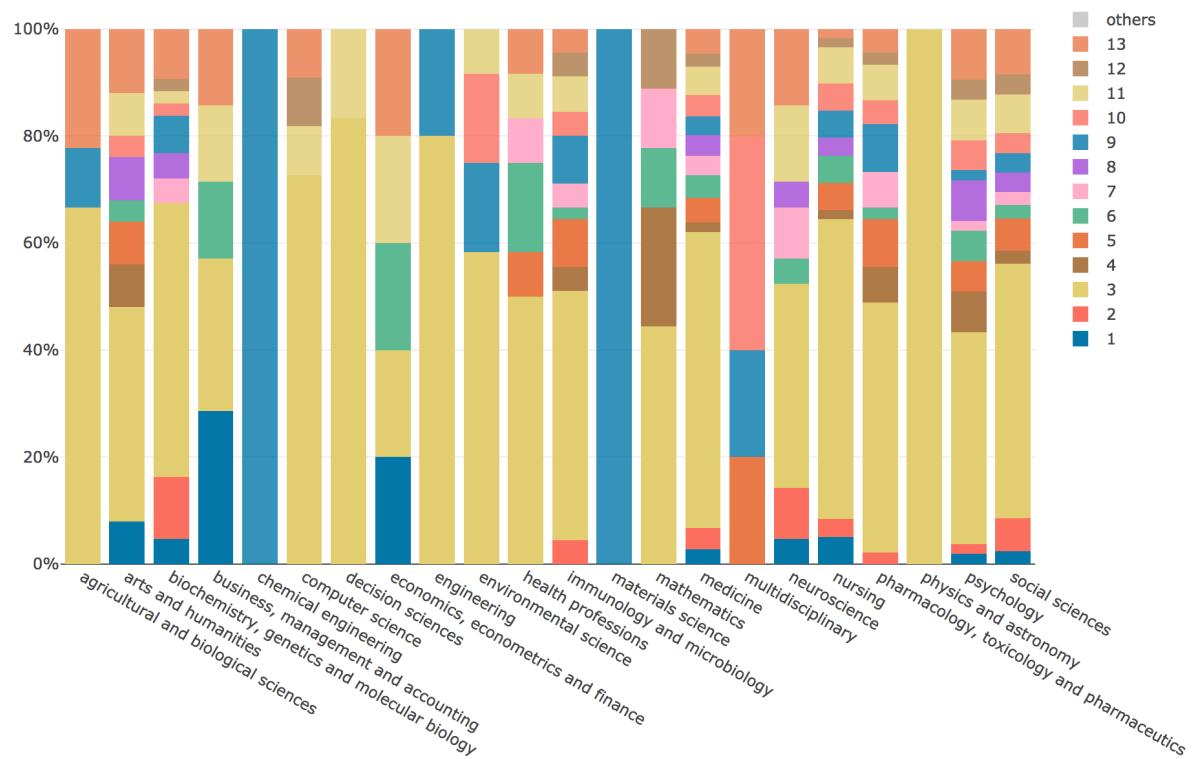


Figure 7. MTMvis built on the topic model obtained from the abstracts of the citing entities, shown against their subject areas.

In-text citations

The total number of in-text citations to WF-PUB-1998 gathered from the 615 citing entities is 870 (1.4 in-text citations per citing entity on average). In Table 3, we list the features we

collected, accompanying them with a brief description (first column) and the corresponding values (second column), i.e. the total number of in-text citations having a value specified for the corresponding feature and, if applicable, a classification of the different possible values.

WF-PUB-1998 in-text citations features	Values
<i>intext_citation.section</i> The kind of section in the citing entity which includes the in-text citation, taken from the list in (Suppe, 1998)	Total: 757 (87%) in-text citations had a value specified. Values: 10 different values: <i>introduction</i> (166), <i>discussion</i> (61), <i>results</i> (28), <i>background</i> (36), <i>conclusions</i> (17), <i>method</i> (15), <i>abstract</i> (5)
<i>intext_citation.context</i> The textual context in the citing entity which includes the in-text citation	Total: all the in-text citations had a value specified.
<i>intext_citation.pointer</i> The string representing the in-text reference pointer (e.g. “[3]”) in the citing entity to the bibliographic reference of WF-PUB-1998	Total: all the in-text citations had a value specified.
<i>intext_citation.intent</i> The citation intent related to the in-text citation in the citing entity, i.e. the author’s reason for citing WF-PUB-1998, taken among the citation functions defined in CiTO	Total: all the in-text citations had a value specified. Values: 17 different values: <i>discusses</i> (226), <i>disputes</i> (114), <i>credits</i> (95), <i>cites for information</i> (90), <i>cites as evidence</i> (74), <i>qualifies</i> (70), <i>describes</i> (60), <i>obtains background from</i> (56), <i>critiques</i> (55), <i>includes excerpt from</i> (8), <i>obtains support from</i> (6), <i>uses data from</i> (5), <i>uses conclusions from</i> (4), <i>ridicules</i> (4), <i>extends</i> (1), <i>updates</i> (1), <i>refutes</i> (1)
<i>intext_citation.sentiment</i> The sentiment , classified as positive/negative/neutral, conveyed by the citation context of an in-text citation	Total: All the in-text citations had a value specified. Values: <i>neutral</i> (549), <i>negative</i> (300), <i>positive</i> (21)

Table 3. The features that directly characterize the in-text citations. The first column lists the features with a brief description, while the second column summarizes the related values we gathered, i.e. the total number and, if applicable, a classification of the different values

Figure 8 shows descriptive statistics of some of the values introduced in Table 3. The sentiment is combined with all the stats displayed (red for negative, yellow for neutral, green for positive). The first column contains the three periods P1-P3 in analysis. The second column shows the distribution per year of the in-text citations, the third column shows the citation intents distribution, and the fourth column shows the in-text citations sections distribution. The sections are classified considering the list proposed in (Suppe, 1998), all the other sections (not part of the list) are grouped under the label “Others”.

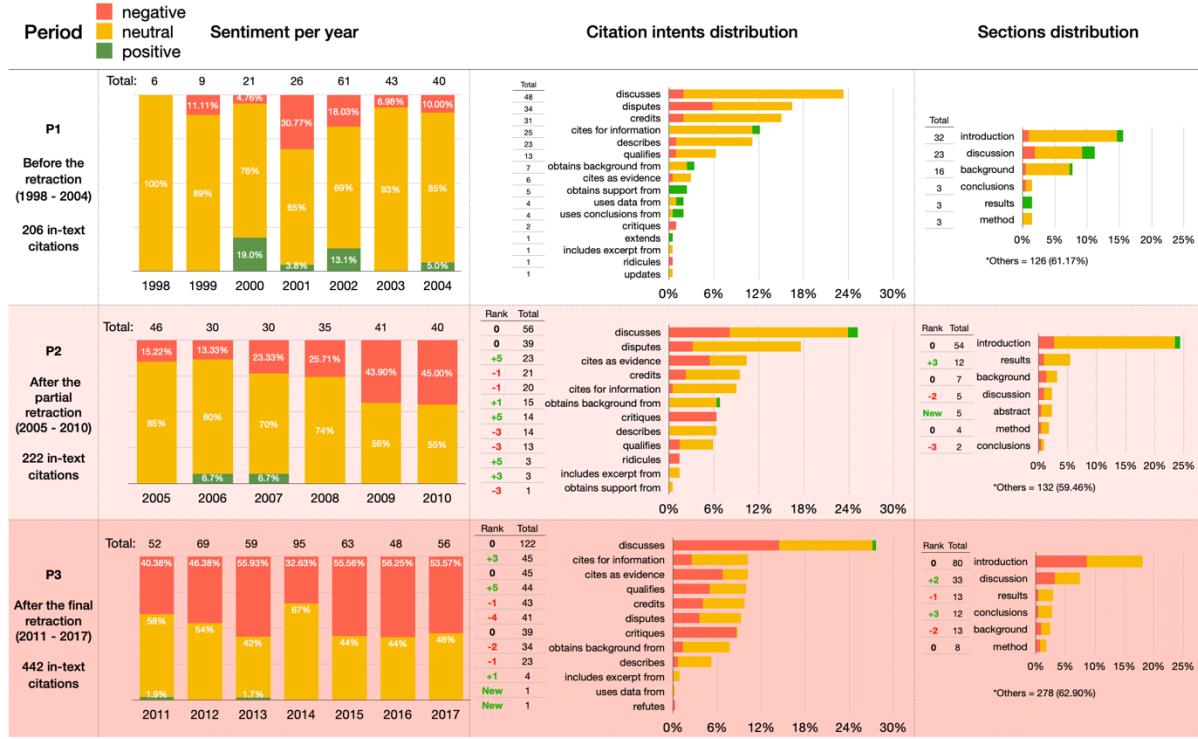


Figure 8. A summary of the in-text citations. All the data are classified under the three sentiments: negative (red), neutral (yellow), and positive (green). The first column contains the periods P1-P3 we considered, the second column shows the distribution per year of the in-text citations, the third column shows the citation intents distribution, and the last column shows the in-text citations sections distribution.

Figure 9 shows the LDAvis of the 22 topics we retrieved using the topic modeling methods described in Section “Topic modeling” by using the citation contexts of in-text citations to WF-PUB-1998. In contrast with the analysis conducted on the abstracts of the citing entities, the 30 most salient terms did not include any term related with the retraction phenomena. The sparsity of the topics in this LDAvis is higher than the one observed with the abstracts and allowed us to spot three different clusters. In particular, we observed two topics with a high prevalence which are also very distant among them (topics 8 and 12). Table 6 (in Appendix) lists all the topics and provides our own interpretation according to their 30 most probable terms.

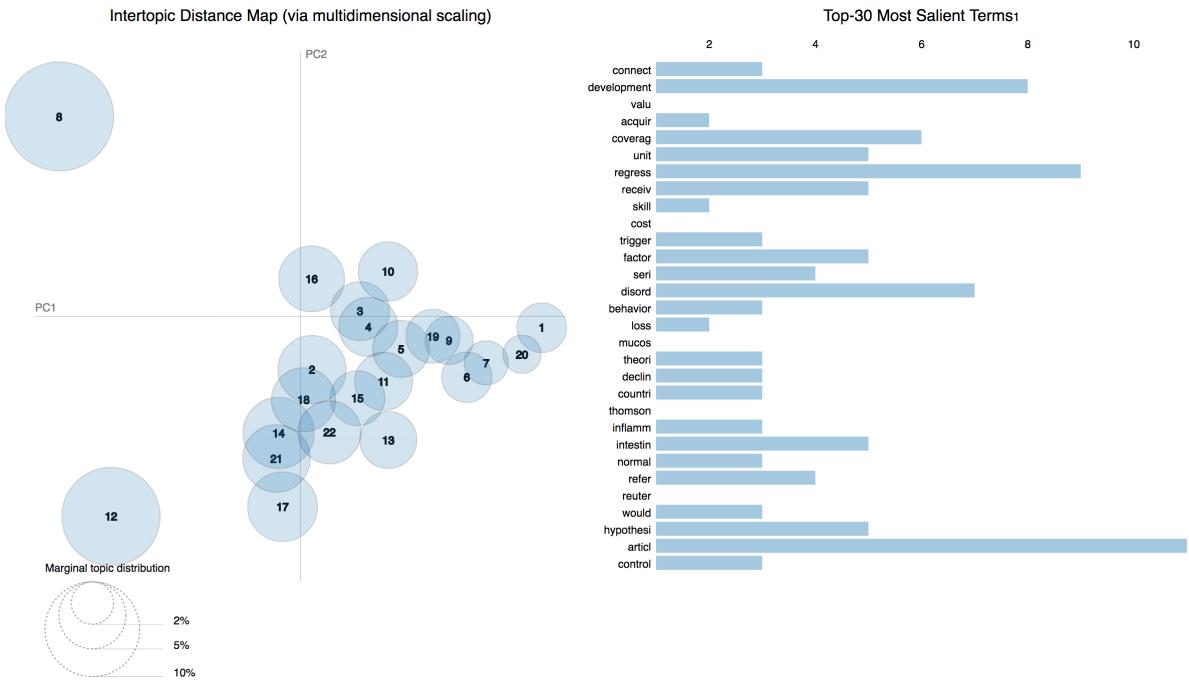


Figure 9. The LDAvis visualization of the topic model created using the citation contexts of the in-text citations contained in the entities citing *WF-PUB-1998*.

The MTMvis visualizations in Figure 10 and Figure 11 refer again to the distribution of the topics over P1-P3 and considering the subject areas of the entities containing the in-text citations analysed. Figure 10 shows that topics 1, 5, 6, 12, and 22 were constantly increasing their percentages throughout P1-P3. Topic 8 and 16, instead, were significantly decreasing along the same period. Topics 2, 4, 10, 13, 19, and 21 had a similar behaviour across P1-P3, having their peak in P2. Contrarily, topics 3, 9, 14, and 18 showed a clear decrease in P2, while in P1 and P3 they showed a similar (and higher) presence.

Figure 11 shows that *medicine*, *social sciences* and *nursing* were the areas of study that included the larger part of the topics identified. Surprisingly, we also had subject areas with a high number of topics which do not concern the medical and social science domains, i.e. *agricultural and biological sciences*, *arts and humanities* and *computer science*.

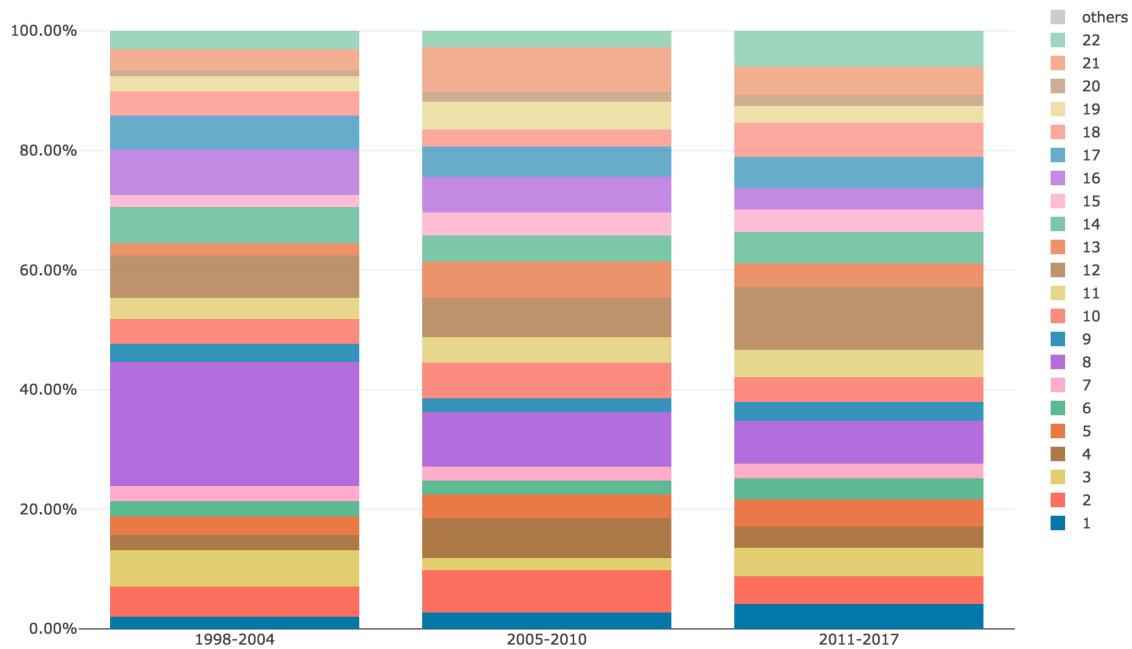


Figure 10. MTMvis created considering the topics extracted from the citation contexts of the in-text citations citing WF-PUB-1998 according to the periods P1-P3.

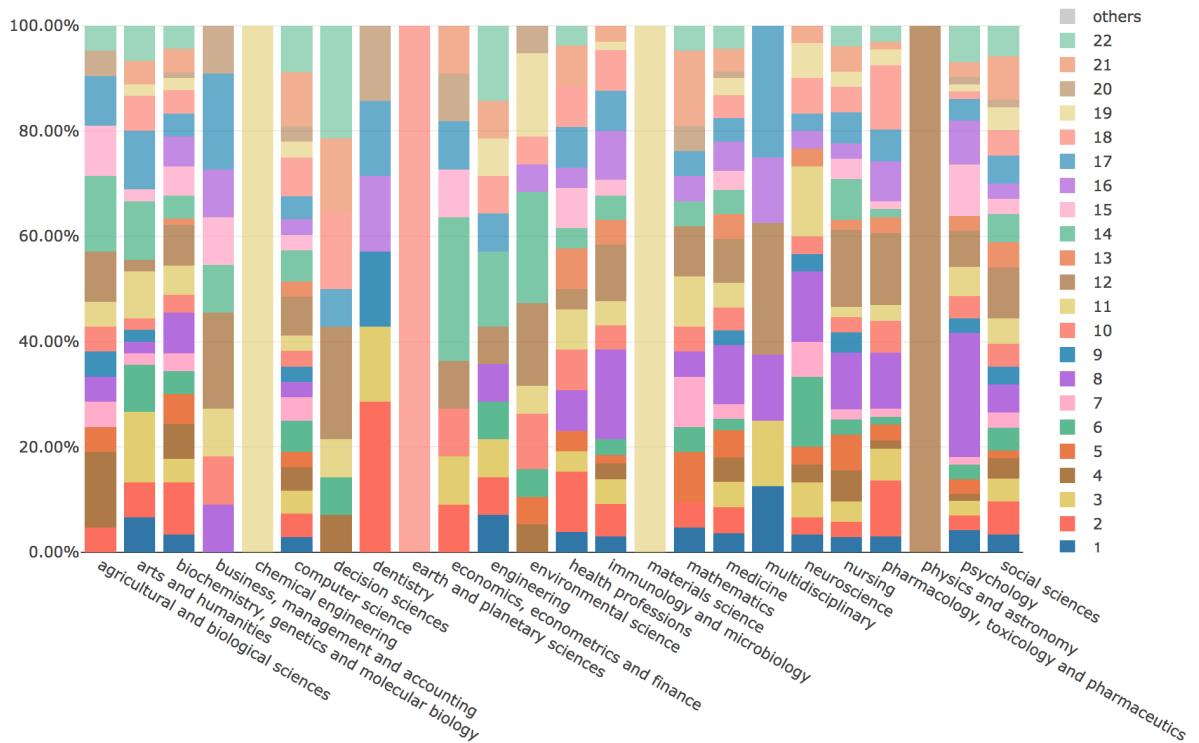


Figure 11. MTMvis created considering the topics extracted from the citation contexts of the in-text citations citing WF-PUB-1998 according to the subject areas of the citing entities.

Discussion

In this section, we discuss the result introduced previously, and we provide insights to answer the two research questions presented in Section “Introduction”. Also, we introduce some limitations of our study and provide suggestions on how to address them in future investigations.

Answering RQ1

From a quantitative point of view, while looking at the subject areas of the citing entities we gathered (see Figure 4), we noticed an increment in the number of areas involved in time. Indeed, the total number of subject areas were 17 in P1 (i.e. before the first partial retraction), while in P2-P3 we counted 22 different subject areas. In addition, in P2-P3 we observed a higher prevalence of non-medical subject areas, i.e. considering the percentage value in P3 with respect to the one in P1, then *social sciences* and *arts and humanities* had increased their percentages, respectively, of 7.81 and 2.21 times more than those observed in P1. On the contrary, considering the same periods (P1 and P3), *medicine* and *nursing* had an inverse trend, since their presence decreased by almost 30% and 40% percent compared with P1, respectively. These figures suggested that the retraction attracted the attention of other subject areas which were not strictly related to the original one of WF-PUB-1998 (medicine).

In addition, we also noticed a continuous increment in the percentage of entities that have explicitly mentioned the retraction of WF-PUB-1998 over the time (see Figure 4). The peak was reached in 2017 (the last year we have considered) with a 61% percentage of entities mentioning the retraction, and we had an important percentage of entities mentioning the retraction even before the full retraction notice (e.g. 25% of entities in 2006). This suggests that the citing entities do not always wait for the full retraction notice before acknowledging the retraction (even if partial), since this can happen before the eventual full retraction. This aspect might be also related to the particular kind of the partial retraction (that was "Concerns/Issues About Results", and "Error in Results and/or Conclusions" in WF-PUB-1998) and with the popularity of the particular case in consideration.

Looking at the retrieved topics in the topic model created using the abstracts of the citing entities, we noticed that topics 1,2 and 5 were those increasing their presence after the partial retraction (i.e. starting from P2). The theematics covered by these topics seemed to refer to discussions on the retraction phenomena (see Table 5 in Appendix) and used a limited number of terms from the medical jargon.

A deeper investigation toward the evolution of topics 1,2 and 5 during P2-P3 on all the subjects areas, showed that topics 2 and 5 had got a significant increment in P3 (11.48% vs 5.15%) while topic 1 has a slighter increment (3.09% vs. 3.28%), as we can see in Figure 12. This might indicate that topic 1 (and the abstracts linked to it) discussed the retraction phenomena similarly over P2-P3. In fact, although topic 1 included words that deal with ethical/social issues (see Table 5 in Appendix), it did not include words strongly related to the retraction, or having a strongly negative sentiment. The citing entities linked to topic 1 had therefore decided to cite WF-PUB-1998 and discuss the case without mentioning the actual retraction of WF-PUB-1998, even after its full retraction (i.e. P3). Figure 13 shows that

topic 1 is mainly related to the *medicine* subject area (excluding the subject areas with limited number of abstracts, e.g. *arts and humanities* with 2 abstracts). This relation between topic 1 and *medicine* is also quite interesting, considering the fact that topic 1 has little engagement with the medical thematics, considering its 30 most probable terms. Thus, part of the entities in the *medicine* subject area discussed the retraction of WF-PUB-1998 in non-medical terms as well.

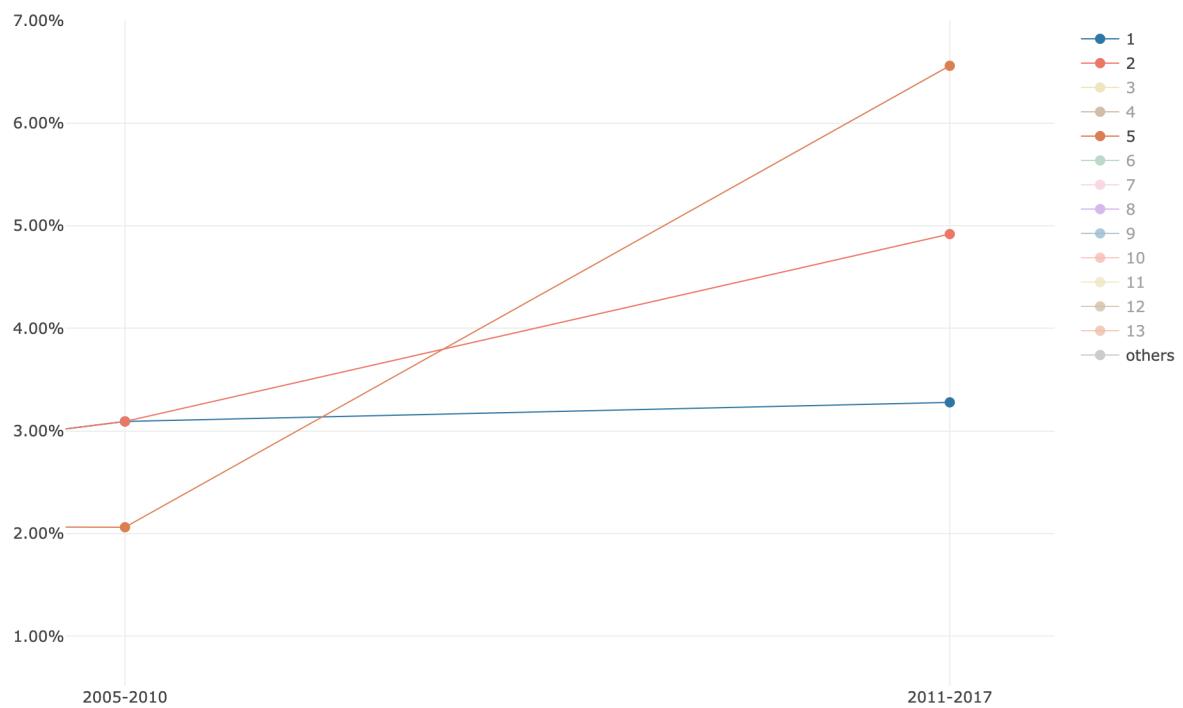


Figure 12. the evolution of topics 1,2 and 5 during P2-P3 on all the subjects areas plotted using MTMvis. MTMvis has been generated on the results of the topic modeling over the abstracts of the citing entities.

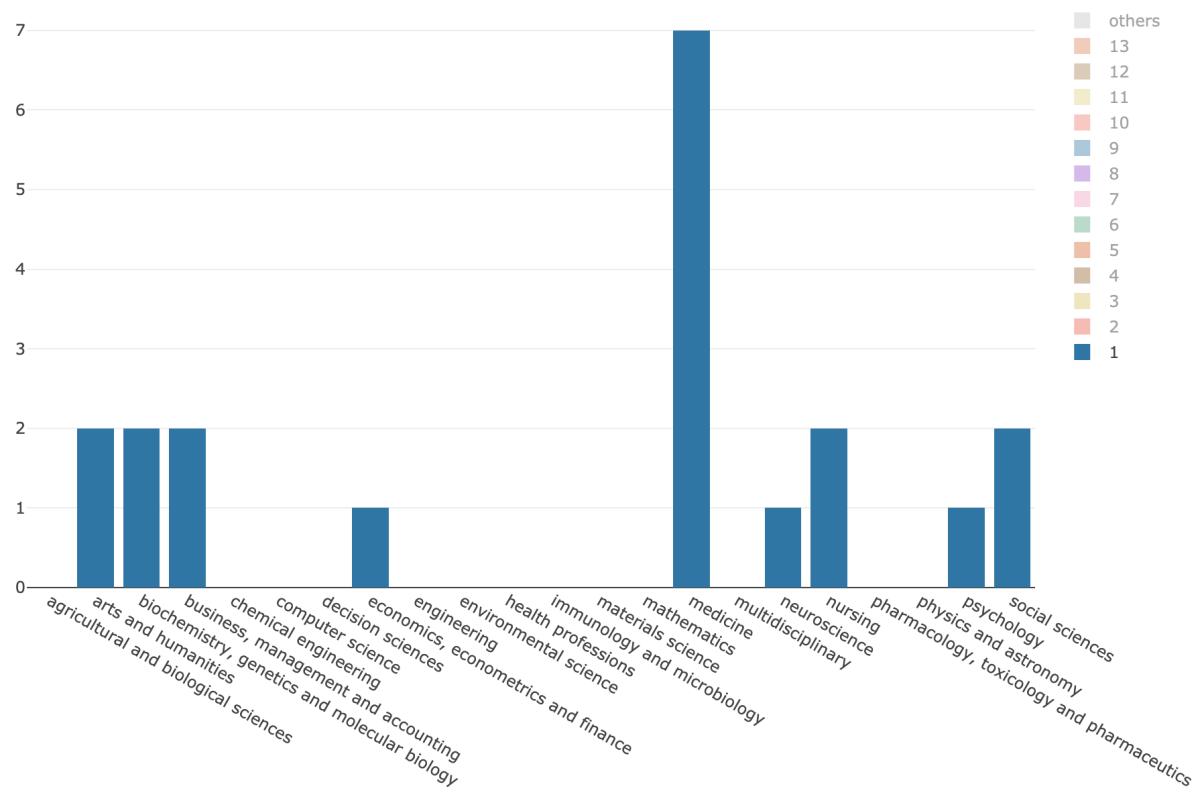


Figure 13. the distribution of topic 1 over all the subject areas during P2-P3 plotted using MTMvis. MTMvis has been generated on the results of the topic modeling over the abstracts of the citing entities

We investigated the distribution of topics 2 and 5 over the subject areas during P2-P3 and checked if such topics were part of the top five ones of each related subject area, as summarised in Figure 14. We can see that topics 2 and 5 were listed in the top five topics of twelve subject areas. Avoiding to consider the subject areas for which we had a small number of abstracts in P2-P3 (e.g. *economics, econometrics and finance* and *multidisciplinary*, both having 1 abstract), we noticed that topics 2 and 5 were highly represented in the *social sciences* with a total percentage of 12% (number of abstract: 12) of all the abstracts in P2-P3. These considerations suggest that topics 2 and 5 were the ones that better represent and characterize the period after the full retraction (i.e. P3), and that *social sciences* is the subject area that dealt the most with the thematics emerged in P3. Contrary to our previous considerations regarding topic 1, in these two topics we found a clear reference to the retraction. The fact that this aspect was manifested in the analysis of the abstracts may indicate that the retraction might have been one of the main subjects discussed in the entities of the abstracts analysed.

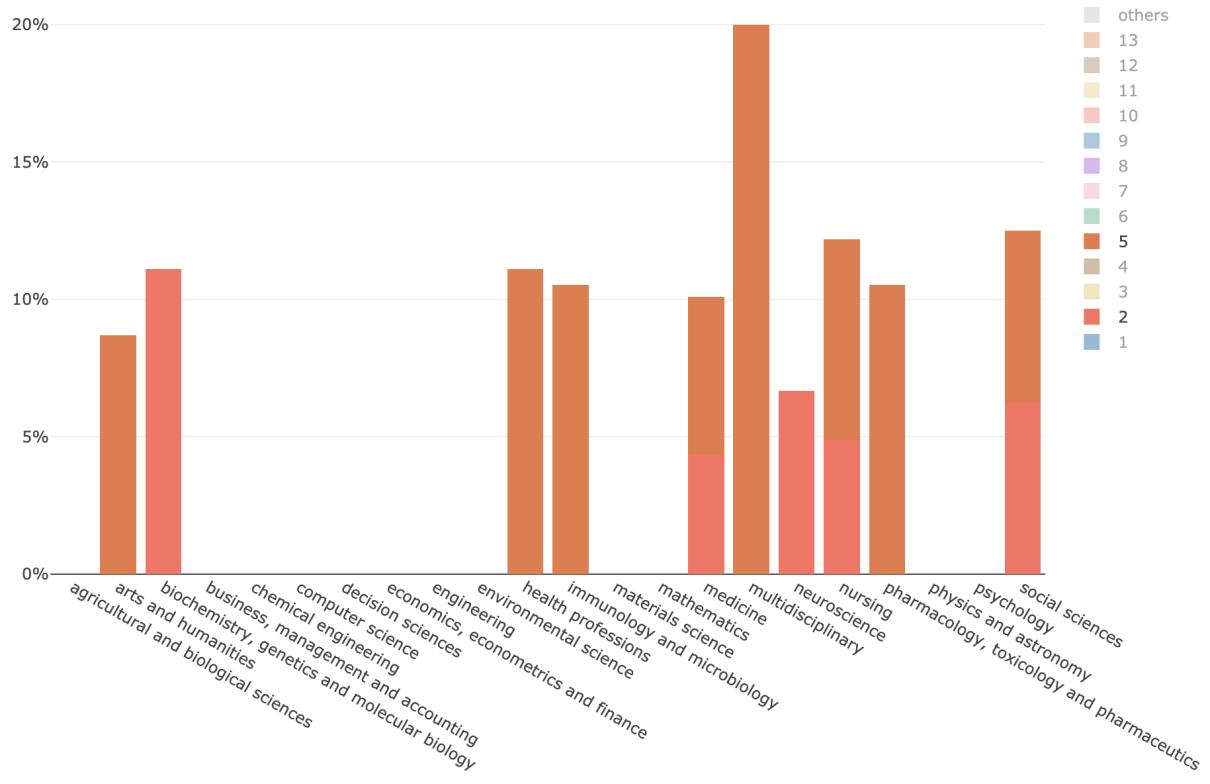


Figure 14. The subject areas of citing entities published in P2-P3 which includes either topic 1,2, or 5 in their top 5 topics.

Answering RQ2

Figure 8 shows that the intended sentiment carried in the citation contexts of the in-text citations referring to WF-PUB-1998 moved to a negative perception over time. However, the retraction of WF-PUB-1998 was not always mentioned in these cases. Indeed, as shown in Figure 8, in 2015 only 32% of the citing entities mentioned the retraction even if the perceived sentiment in the same year is either negative (for 55.56% of in-text citations) or neutral (for 44% of in-text citations).

The distribution of the citation intents annotated in the in-text citations during P1-P3 showed an increment in the use of general purpose intents such as *discusses* and *cites for information*. This could be highly related with increasing popularity of the retraction of WF-PUB-1998 in the non-medical subject areas (as already stated in the previous section). Probably, the entities that are part of the non-medical subject areas cited WF-PUB-1998 from a general perspective, without recalling strictly medical details in their text, which are out of the scope of their research domains.

As shown in Figure 15. The set of intents uses *conclusions from*, *updates*, *extends*, *uses data from*, and *obtains support from* decreased starting from P2, probably due to a lesser use of the data and conclusions contained in WF-PUB-1998 after its retraction. Other citation intents, instead, showed a clear increment of their use along the three periods. For instance, the use of *critiques* seemed to be related somehow with the increment of the negative sentiment overall. Instead, *credits* had an important drop. In this case, the citing entities published before the partial retraction of WF-PUB-1998 used it mostly in a neutral way to

credit Wakefield et al. for their findings. However, in P2-P3, beside the overall drop, *credits* had a higher percentage of negative citations. This last aspect was also noticed in the intent *cites as evidence*, although its overall usage has increased in time. However, if before the retraction, *cites for evidence* was used neutrally to refer to WF-PUB-1998 to support some statements or conclusions in the citing entities, after the retraction it was actually used to highlight WF-PUB-1998 as a negative scientific example due to its retraction and, more generally, of a faulty science.

In Figure 16 we investigated the sections of the in-text citations marked as *credits* and *cites as evidence*. On the one hand, the *credits* citations were mostly distributed on descriptive sections – i.e. *introduction*, *discussion*, and *background* – during all the three periods. On the other hand, the *cites as evidence* citations appeared also in technical sections – i.e. *results*, and *method*. The sections distribution in P3 for both *credits* and *cites as evidence* followed the overall distribution we presented in Figure 8, such that the in-text citations have been concentrated in few sections mostly of descriptive type - i.e. *introduction*, *discussion*.

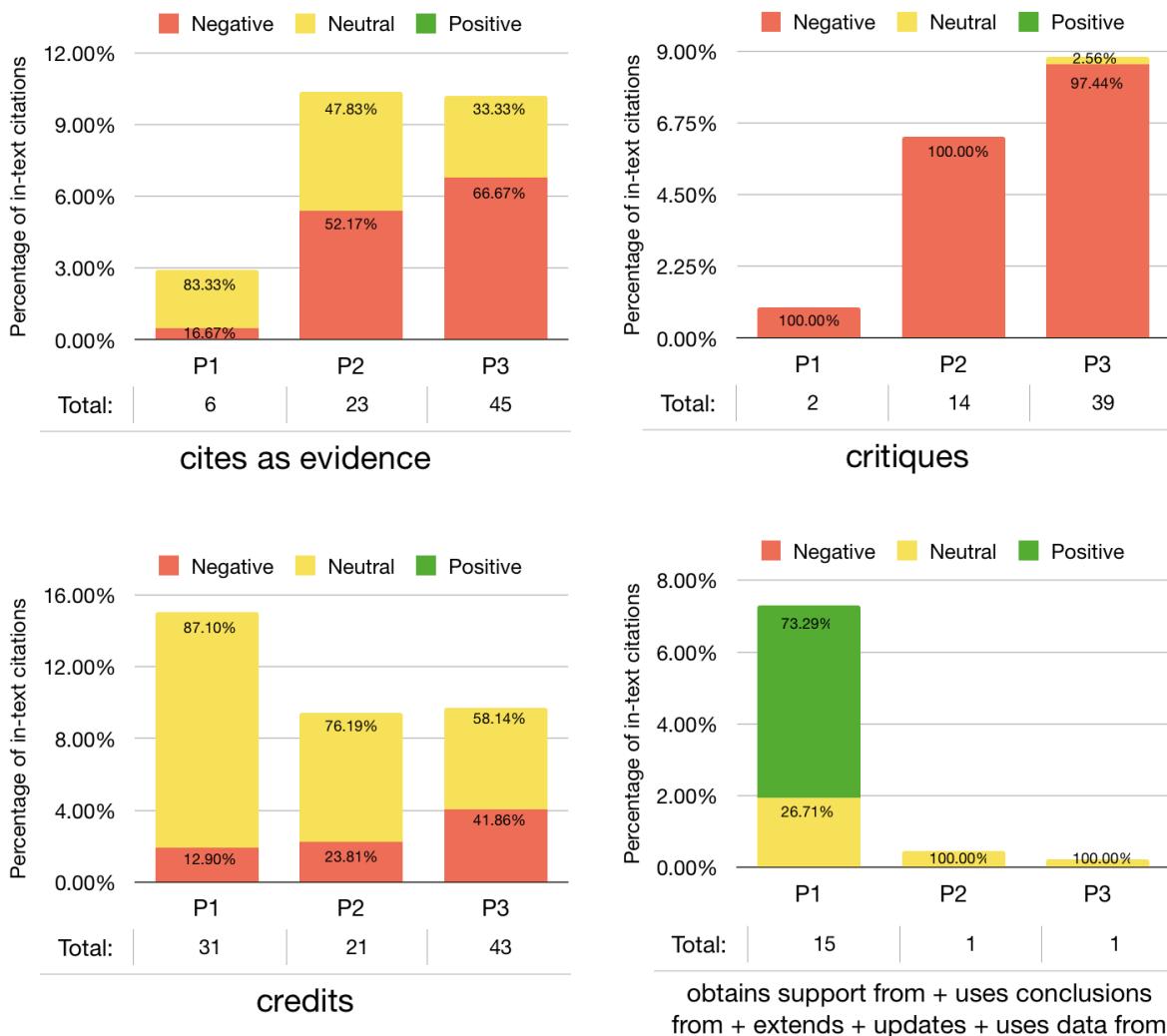


Figure 15. The four graphs illustrate the way the use of citation intents changed over time and according to their perceived sentiment.

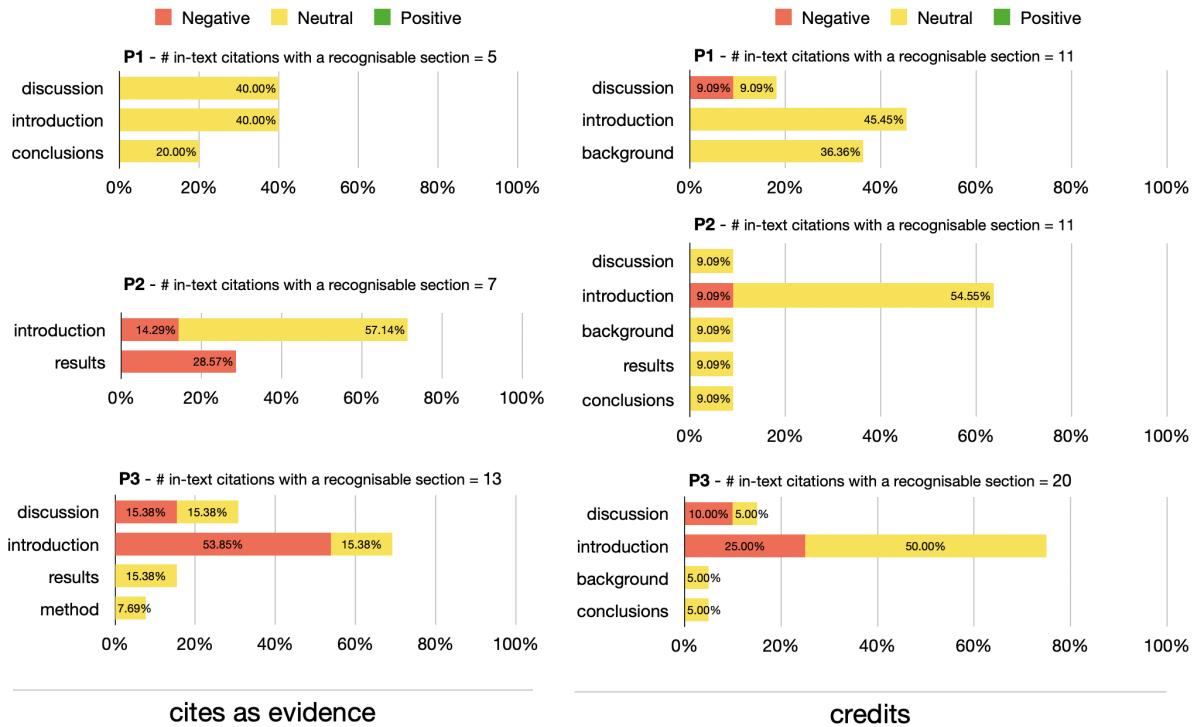


Figure 16. the *cites as evidence* and *credits* citation intents distribution among the sections (the recognizable ones) and during the three periods (i.e. P1-P3).

We analysed the twenty-two topics we obtained considering the topic model created using the citation contexts of the in-text citations referring to WF-PUB-1998. In particular, as shown in Figure 17, we focussed on:

1. the topics for which we observed an increasing use over time;
2. the topics which had a huge increment in their use in P3;
3. the topics which had a constant decrement in their use over time.

The topics that increased over P1-P3 (i.e. topics 1, 5, and 11) included a few medical terms and seemed to refer to the controversy of the retraction of WF-PUB-1998 from a mathematical and statistical perspective. A second group of topics (i.e. topics 12, 18 and 22) seemed to refer to WF-PUB-1998 as an example of faulty science, which was acknowledged clearly in P3. The drastic change of these topics in P3 is very significant. Indeed, all the three topics (as shown Table 6 in Appendix) mention the word “retraction” (and its derivatives) along with other words with a strong negative connotation. In other words, it seems that the authors waited the full retraction notice before marking their negative impressions toward WF-PUB-1998 – 19.8% of the citations in P3 are part of this group of topics.

A similar behaviour could be noticed also in the citations coming from medical subject areas, since 22.97% and 30.61% of the citations in P3 are coming from *medicine* and *nursing* articles, respectively. This suggests that also the entities close to the domain of the retracted article did not hesitate to judge a retracted work done by their colleagues.

The last group of topics (i.e. topics 8 and 16) were mainly related to the medical domain, and included some medical backgrounds of WF-PUB-1998. The fact that these topics had a clear

decrement over time suggests that the most recent citing works provided partial and limited acknowledgement of the conclusions and medical arguments in WF-PUB-1998.

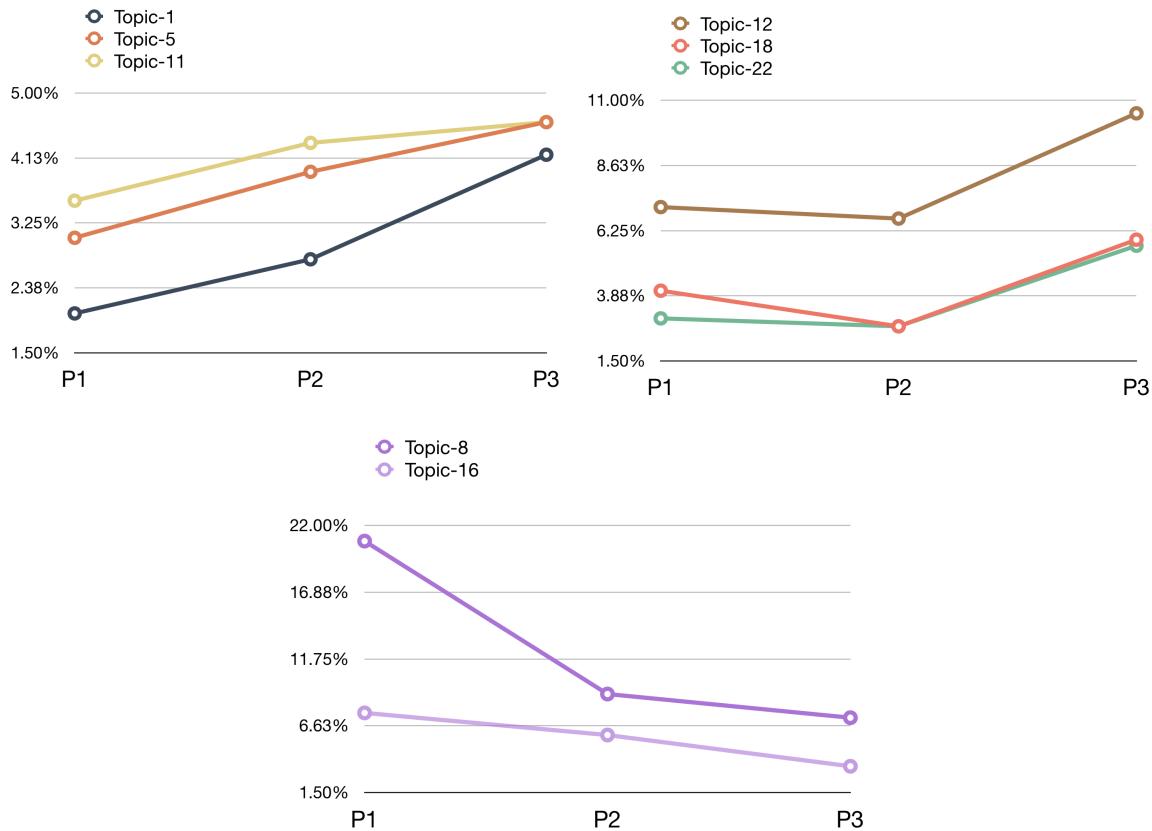


Figure 17. The evolution of the presence of three groups of topics defined from the citation contexts of the in-text citations to WF-PUB-1998 over time.

In Figure 18, we show the topics that either increased (left panel) or decreased (right panel) their presence over time considering only the citation contexts of the citing entities belonging to the *medicine* subject area. Some of the topics shown in Figure 18 are also included in Figure 17, although there is an important difference: topic 15 (that concerned the conclusions of WF-PUB-1998 and the controversies arising from it) is not listed in Figure 17, even if it seemed relevant when we focus only on the *medicine* subject area. We had a similar situation also with the topics decreasing over time. Indeed, topic 7 (which summarizes WF-PUB-1998 and medical conclusions) is not highlighted in Figure 17 as well.

This scenario suggests that the citing entities in the *medicine* subject area included additional prominent topics when discussing WF-PUB-1998. More precisely, after its final retraction (i.e. P3), part of the entities addressed the retraction through a discussion using medical terms. The decreasing relevance of topic 7 indicates that also the entities part of the same subject area of the retracted article (i.e. *medicine*) addressed less the subjects treated in WF-PUB-1998, and rather focused on citing and discussing the retraction of WF-PUB-1998 without deepening in its medical theatics.

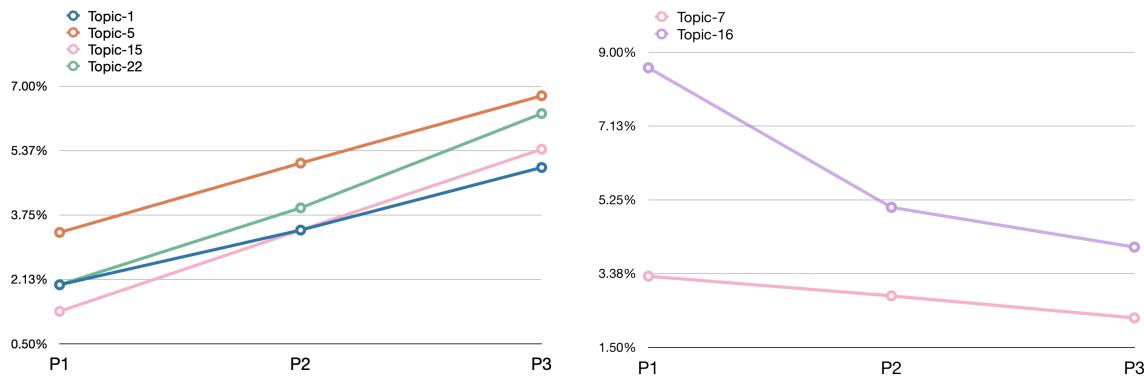


Figure 18. The evolving (left) and declining (right) topics popularity of the in-text citation topic model, considering only the “medicine” area of study

Limitations of our study and future suggestions

Our findings and observations provide additional insights on the retraction of WF-PUB-1998 and how it has been perceived by the scientific community. However, we are aware of particular limitations that may have affected the findings and the interpretations we made throughout this study. In this section we list the methodological limitations and we compare our outcomes with previous related works on the same topic.

First, we used the data in COCI to gather all the citations to WF-PUB-1998 used in our study. Since COCI contains citations between entities included in Crossref when they are both identified by DOIs, we did not include in our analysis all the citations that involved entities with no DOIs. Also, we missed the citations to WF-PUB-1998 from articles published by some publishers, such as Elsevier, that did not share openly their reference lists via Crossref in 2018 – and that, thus, were not available in the COCI dump, i.e. the November 2018 release (OpenCitations, 2018).

For a few citing entities (i.e. 22) involved in the citations we gathered, we could not retrieve their full text due to commercial paywalls, preventing us to analyse the citation contexts and in-text citations they defined. Thus, we excluded these citing entities, and their related citations, from our analysis.

While working on similar problems, the data we gathered in our study are slightly different from those used in (Suelzer et al., 2019), which introduces an analysis of WF-PUB-1998, as anticipated in the introduction. In particular, Suelzer et al. collected 1,211 articles from a Web of Science Core Collection search performed in March 2019, while we collected citations coming from 615 articles in total. Similarities and differences between our study and Suelzer et al.’s one are introduced in Table 4.

Method and results in (Suelzer et al., 2019) compared with our study		
Feature	Method	Results
Mentions of the retraction	Differences: 1) The strategy to select citation context of each in-text citation is not specified. 2) The annotators read an arbitrary citation	Differences: 1) Compared to our study, a higher number of entities mention the retraction (since the citation

	<p>context and annotate this value</p> <p>Similarities: 1) A citing entity mentions the retraction, only if the word “retraction” (and its derivatives) is used.</p>	<p>context taken into is probably larger)</p> <p>Similarities: 1) The trend in P2-P3 is similar and shows a continuous increment in the number of entities mentioning the retraction over time. 2) 2009 was the year with the lowest number of entities mentioning the retraction.</p>
Citation Intent	<p>Differences: 1) The citations are characterized into 8 different categories, following the definitions in (Bornmann & Daniel, 2008).</p> <p>Similarities: 1) The annotation was performed following a set of rules which guided and helped the annotator</p>	<p>Differences: 1) The plotted results combine in the same dimension both the sentiment and the citation function thus it is difficult to compare them with our results</p>
Author sentiment	<p>Differences: 1) Not part of the analysed features 2) This information is embedded in the citation intent value</p>	
Citations sections	<p>Differences: 1) Not part of the analysed features 2) Some articles are cited as examples of citations accomplished from the <i>introduction</i> and <i>discussion</i> sections.</p>	
Text analysis (i.e. topic modeling)	<p>Differences: 1) Not part of the analysed features 2) They give a brief summary regarding their observations of some examples that have cited WF-PUB-1998</p>	

Table 4. A summary of the differences and similarities between our study and (Suelzer et al., 2019).

Another important aspect of our study is the manual annotation of citation intents. Although the annotation has been done carefully by following a specific methodology, it was based on a subjective interpretation of the text and, thus, may differ from the original citation intent that the authors of the citations to WF-PUB-1998 had in mind.

In addition to the limits regarding our methods and findings, there are also other aspects that this work did not address compared with the past approaches. In particular, we would like to work, in future developments of this research, in the generation of a citation network starting from either our seed retracted article or from its citing entities, as suggested by van der Vet and Nijveen (2016) who proved the importance of such analysis, since it might enlighten us on the negative/positive outcomes of the propagation of retracted research results.

Conclusions

In this paper we have presented the outcomes of a citation analysis of a highly cited and popular retracted article – i.e. (Wakefield et al., 1998) (i.e. WF-PUB-1998). We have applied a quantitative and qualitative analysis of the citations that cited WF-PUB-1998 and we

clustered them into three periods: (P1) before the WF-PUB-1998 partial retraction, (P2) after the partial retraction and before its full retraction, and (P3) after its full retraction. The main purpose of this work was to understand the retraction phenomenon and how it was perceived from the scientific community when referring to retracted articles in their own work. WF-PUB-1998 is a popular example of a retracted article that was highly cited by other works over time (before and after the retraction notes), therefore we considered it as a perfect example to analyse. We approached our general goal through the definition of two research questions aiming at analysing possible evolution, before and after the retraction, of the research topics addressed by the articles citing WF-PUB-1998 and the main characteristics of such citations. To answer these questions we have defined a methodology which allowed us to gather data, to automatically process the textual information retrieved (abstract and citation context) to extract topics and, thus, to address the research questions.

Our results have been presented in two categories according to the entities that were analysed: citing entities and their in-text citations. We first showed a quantitative overview of the features we have collected, and then we discussed the outcomes of the topic models obtained. Finally, in Section “Discussion”, we discussed all the evidence we have collected to answer the research questions. In particular, we observed that (a) the citing entities generally did not wait for a full retraction notice before acknowledging the retraction of the cited article, (b) the *social sciences* subject area is the one that dealt the most with the retraction of WF-PUB-1998, (c) the authors of the citing articles used to talk about WF-PUB-1998, after its retraction, from a general perspective without recalling strictly medical details in their text.

Finally, we have also discussed the limits of our approach from a methodological point of view, and compared our methods and results with the ones of (Suelzer et al., 2019). The bigger difference has regarded the additional features we have considered in our analysis – i.e. the author sentiment, the citations sections, and the topic modeling analysis. Many of our findings confirmed also the results of Suelzer et al.’s work.

References

- Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy*, 46(9), 1552–1569.
<https://doi.org/10.1016/j.respol.2017.07.003>
- Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: A case study. *Scientometrics*, 113(1), 547–565. <https://doi.org/10.1007/s11192-017-2242-0>
- Bengfort, B., Bilbro, R., & Ojeda, T. (2018). *Applied text analysis with Python: Enabling language-aware data products with machine learning* (First edition). O'Reilly Media, Inc.
- Bornemann-Cimenti, H., Szilagyi, I. S., & Sandner-Kiesling, A. (2016). Perpetuation of

Retracted Publications Using the Example of the Scott S. Reuben Case: Incidences, Reasons and Possible Improvements. *Science and Engineering Ethics*, 22(4), 1063–1072. <https://doi.org/10.1007/s11948-015-9680-y>

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.

<https://doi.org/10.1108/00220410810844150>

Chen, C., & Leydesdorff, L. (2014). Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology*, 65(2), 334–351.

<https://doi.org/10.1002/asi.22968>

Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, 74.

<https://doi.org/10.1145/2254556.2254572>

Collier, R. (2011). Shedding light on retractions. *Canadian Medical Association Journal*, 183(7), E385–E386. <https://doi.org/10.1503/cmaj.109-3827>

Heibi, I., & Peroni, S. (2020). *A methodology for gathering and annotating the raw-data/characteristics of the documents citing a retracted article v1* (*protocols.io.bdc4i2yw*) [Data set]. <https://doi.org/10.17504/protocols.io.bdc4i2yw>

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, 121(2), 1213–1228.

<https://doi.org/10.1007/s11192-019-03217-6>

Heibi, Ivan, & Peroni, Silvio. (2020). *Methodology data of “A qualitative and quantitative citation analysis toward retracted articles: A case of study”* [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.4323221>

Jan, R., Bano, S., Mehraj, M., & others. (2018). Context Analysis of Top Seven Retracted Articles: Should Retraction Watch Revisit the List? *Context*. <https://digitalcommons.unl.edu/libphilprac/2016/>

- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211.
<https://doi.org/10.1007/s11042-018-6894-4>
- Lu, S. F., Jin, G. Z., Uzzi, B., & Jones, B. (2013). The Retraction Penalty: Evidence from the Web of Science. *Scientific Reports*, 3(1), 3146. <https://doi.org/10.1038/srep03146>
- Luwel, M., van Eck, N. J., & van Leeuwen, T. N. (2019). *The Schön case: Analyzing in-text citations to papers before and after retraction* [Preprint]. SocArXiv.
<https://doi.org/10.31235/osf.io/c6mvs>
- May, C., Cotterell, R., & Van Durme, B. (2019). An Analysis of Lemmatization on Topic Models of Morphologically Rich Language. *ArXiv:1608.03995 [Cs]*.
<http://arxiv.org/abs/1608.03995>
- Meyer, C. A. (2011). Distinguishing published scholarly content with CrossMark. *Learned Publishing*, 24(2), 87–93. <https://doi.org/10.1087/20110202>
- Mongeon, P., & Larivière, V. (2016). Costly collaborations: The impact of scientific fraud on co-authors' careers: Costly Collaborations: The Impact of Scientific Fraud on Co-Authors' Careers. *Journal of the Association for Information Science and Technology*, 67(3), 535–542. <https://doi.org/10.1002/asi.23421>
- Mott, A., Fairhurst, C., & Torgerson, D. (2019). Assessing the impact of retraction on the citation of randomized controlled trial reports: An interrupted time-series analysis. *Journal of Health Services Research & Policy*, 24(1), 44–51.
<https://doi.org/10.1177/1355819618797965>
- Moylan, E. C., & Kowalcuk, M. K. (2016). Why articles are retracted: A retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open*, 6(11), e012047.
<https://doi.org/10.1136/bmjopen-2016-012047>
- OpenCitations. (2018). COCI CSV dataset of all the citation data (p. 11568165723 Bytes) [Data set]. figshare. <https://doi.org/10.6084/M9.FIGSHARE.6741422.V3>
- Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic

resources and citations. *Journal of Web Semantics*, 17, 33–43.

<https://doi.org/10.1016/j.websem.2012.08.001>

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic Modeling as a Strategy of Inquiry

in Organizational Research: A Tutorial With an Application Example on

Organizational Culture. *Organizational Research Methods*, 22(4), 941–968.

<https://doi.org/10.1177/1094428118773858>

Shuai, X., Rollins, J., Moulinier, I., Custis, T., Edmunds, M., & Schilder, F. (2017). A

Multidimensional Investigation of the Effects of Publication Retraction on Scholarly

Impact. *Journal of the Association for Information Science and Technology*, 68(9),

2225–2236. <https://doi.org/10.1002/asi.23826>

Sievert, C., & Shirley, K. E. (2014). *LDAvis: A method for visualizing and interpreting topics*.

<https://doi.org/10.13140/2.1.1394.3043>

Suelzer, E. M., Deal, J., Hanus, K. L., Ruggeri, B., Sieracki, R., & Witkowski, E. (2019).

Assessment of Citations of the Retracted Article by Wakefield et al With Fraudulent

Claims of an Association Between Vaccination and Autism. *JAMA Network Open*,

2(11), e1915552. <https://doi.org/10.1001/jamanetworkopen.2019.15552>

Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, 65(3), 381–405.

Teixeira da Silva, J. A., & Dobránszki, J. (2017). Highly cited retracted papers.

Scientometrics, 110(3), 1653–1661. <https://doi.org/10.1007/s11192-016-2227-4>

Truica, C.-O., Radulescu, F., & Boicea, A. (2016). Comparing Different Term Weighting

Schemas for Topic Modeling. *2016 18th International Symposium on Symbolic and*

Numeric Algorithms for Scientific Computing (SYNASC), 307–310.

<https://doi.org/10.1109/SYNASC.2016.055>

van der Vet, P. E., & Nijveen, H. (2016). Propagation of errors in citation networks: A study

involving the entire citation network of a widely cited paper published in, and later

retracted from, the journal Nature. *Research Integrity and Peer Review*, 1(1), 3.

<https://doi.org/10.1186/s41073-016-0008-5>

Wakefield, A., Murch, S., Anthony, A., Linnell, J., Casson, D., Malik, M., Berelowitz, M.,

Dhillon, A., Thomson, M., Harvey, P., Valentine, A., Davies, S., & Walker-Smith, J. (1998). RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103), 637–641.
[https://doi.org/10.1016/S0140-6736\(97\)11096-0](https://doi.org/10.1016/S0140-6736(97)11096-0)

Appendix

Topic (proportion)	Terms (the 30 most probable terms)	Interpretation
1 (3.5%)	<i>bias, epidem, philosoph, cell, experi, consum, behavior, protect, even, illich, acetaminophen, scientist, scienc, oxid, call, conserv, long, public, scientif, phone, campaign, occur, experiment, ethic, feed, among, politician, reject, dissent, insignific</i>	Close to the social studies domain, might take in consideration ethical thematics.
2 (4.2%)	<i>retract, symptom, disabl, student, scienc, perspect, expertis, misconduct, forens, paper, probabl, studi, mandat, cite, diseas, treat, provid, research, replic, vaccin, tripl, librari, qualif, adjuv, might, take, younger, case, media, wakefield</i>	Includes terms related to the retraction phenomena (study-domain independent), it includes terms used in scientometric analysis.
3 (46%)	<i>vaccin, parent, children, health, inform, immun, public, decis, review, safeti, studi, measl, evid, articl, research, risk, disord, practic, autist, issu, factor, concern, diseas, report, increas, import, relat, child, effect, base</i>	This is by far the larger topic out of the 13, it contains common terms, highly frequent in the corpus, and close to the WF-PUB-1998 thematics.
4 (3.6%)	<i>access, nurs, bowel, knowledg, hepat, polici, global, mobil, portfolio, immunis, ask, biblic, newspap, pharmaceut, huge, visitor, time, citizen, percept, symptom, organ, internet, take, model, statist, carer, epoch, golden, scientif, cognit</i>	Includes terms from the Medical and pharmaceutical field.
5 (5.1%)	<i>vaccin, regress, anti, myth, movement, countri, syndrom, social, incom, diagnost, affect, determin, right, iron, hcws, children, overview, diseas, court, peopl, parent, million, routin, acupunctur, danger, mortal, immun, claim, degrad, intervent</i>	Some terms are out of the medical field of study, and might indicate a possible discussion.
6 (3.8%)	<i>statist, cultur, describ, infant, citat, exempl, articl, symptom, case, american, journal, metabol, combin, parent, disagr, doctor, abl, associ, construction, bordetella, basi, advers, illustr, gliadin, illusori, literatur, indic, unit, eat, rather</i>	A high number of terms are related to the scientometric field of study. All the terms are objective and don't indicate an opinion or a discussion.
7 (4.1%)	<i>biomark, altmetr, nan, behavior, disord, occur, well, postpon, answer, herd, context, fear, genet, graphic, appeal,</i>	Includes terms from the biology, pharmacology and genetics field of study.

	<i>interquartil, order, vaccin, gene, sensori, evalu, modul, geneticist, chapter, lymphocyt, abstain, putat, approach, protect, homeopathi</i>	Might also indicate a statistical analysis along with an open discussion.
8 (4.9%)	<i>vaccin, balanc, symptom, risk, reaction, link, subgroup, mump, regress, sphere, aefi, opioid, public, prevent, record, case, food, diseas, chronic, media, claim, allerg, week, resid, advers, children, estim, strain, cobalamin, associ</i>	A large part of the terms are close to WF-PUB-1998 treated thematic. Mostly from the medical field of study.
9 (5.1%)	<i>fraud, diseas, narrat, vaccin, complic, health, controversi, comorbid, measl, polici, coliti, bowel, neurolog, travel, inflammatori, movement, trust, ocean, research, retract, attribut, percept, public, futur, preserv, caus, ulcer, case, medic, problem</i>	Concern the retraction phenomena, followed with some medical expressions. It also includes strong terms such as "fraud".
10 (4.9%)	<i>vaccin, immun, misconduct, polici, patholog, retract, aefi, result, children, research, disord, qualit, report, caus, development, advanc, case, chang, internet, record, expos, mold, infecti, program, vaer, actor, live, sinc, mani, appli</i>	General terms related to WF-PUB-1998 treated thematic, some are correlated with a discussion around the retraction phenomena.
11 (5.6%)	<i>vaccin, health, immunis, engag, communic, reason, disord, virus, diagnost, messag, coverag, examin, make, chang, accept, client, diseas, measl, development, research, consid, resist, peopl, public, evid, observ, recent, imag, effect, pervas</i>	Terms from the medical study field related to WF-PUB-1998 treated thematic.
12 (3.3%)	<i>uncertanti, boy, scientif, gynecologist, debat, twitter, semant, ongo, intent, vaccin, disturb, variabl, messag, rhetor, liabil, frame, reddit, percept, content, sourc, gfcf, produc, paediatr, rais, pyridox, guilt, fact, advic, link, chang</i>	Includes terms close to the computer science lexicon and from the pediatric field of study.
13 (5.9%)	<i>vaccin, incid, scientif, erad, measl, frame, viral, literatur, enceph, diseas, controversi, mother, workshop, differ, propos, expert, infect, increas, evalu, genet, dramat, recent, coalit, frequent, communic, current, link, programm, polio, scienc</i>	Includes a large number of general terms from different fields of study, part of them are correlated with WF-PUB-1998 thematic.

Table 5. The 13 topics generated from the abstracts topic model. For each topic (row) we mention its proportion percentage in the corpus (column 1), and the 30 most relevant terms (column 2), and we give our interpretation of it (column 3).

Topic (proportion)	Terms (the 30 most probable terms)	Interpretation
1 (2.7%)	<i>valu, reuter, thomson, worth, retract, would, impact, time, mean, paper, immun, comparison, figur, lancet, base, mening, cerebr, associ, identifi, greater, roach, regress, senior, later, europ, vitamin, viral, assign, consist, campaign</i>	Includes few terms from the medical domain. Might talk about WF-RET-CASE, and from a statistical/mathematical perspective. Some terms include general info about the paper (metadata).

2 (5.2%)	<i>articl, control, associ, public, signific, affect, research, biopsi, scientif, follow, controversi, natur, patient, case, decad, preserv, evid, report, vaccin, use, multipl, clinic, subsequ, differ, first, cell, follicl, symptom, concern, studi</i>	General terms which summarize what WF-PUB-1998 talks about.
3 (3.9%)	<i>articl, case, colon, development, lancet, enterocol, report, assert, associ, public, vaccin, disord, follow, signific, diagnosi, base, group, mumpsrubella, treatment, controversi, parent, scientif, evid, result, symptom, studi, publish, link, reaction, unknown</i>	General terms which summarize what WF-PUB-1998 talks about.
4 (3.9%)	<i>articl, immun, dose, claim, lancet, requir, discredit, declin, find, infect, diseas, februari, colleagu, herd, coverag, respons, measl, report, first, sever, vaccin, regress, month, andrew, second, mump, research, countri, detail, parent</i>	General terms which summarize what WF-PUB-1998 talks about. This topic might also outline other information related to the paper.
5 (3.6%)	<i>vaccin, link, research, suggest, report, appear, measl, focus, scientif, develop, univers, reject, possibl, studi, associ, mump, investig, though, fund, even, continu, around, wakefielddet, newspap, result, doubt, relat, signific, high, evid</i>	Discusses the controversy around WF-RET-CASE.
6 (2.8%)	<i>design, studi, expert, consider, parent, bias, receiv, requir, risk, messag, control, occur, point, factor, time, start, howev, report, third, call, know, vaccin, connect, major, qualiti, media, research, school, best, mump</i>	Does not include any medical term, it rather focuses on other related aspects concerned with WF-PUB-1998
7 (2.2%)	<i>diseas, caus, characterist, process, read, inflamm, alarm, report, assert, without, show, ileocolon, bowel, declin, safeti, student, first, exempl, intestin, media, young, detail, autist, adult, possibl, scientif, studi, even, control, wide</i>	General terms which summarize what WF-PUB-1998 talks about. Many terms are related to its medical background.
8 (13.2%)	<i>regress, development, increas, hypothes, causal, vaccin, link, bowel, disord, case, report, measl, problem, symptom, author, associ, system, immun, mump, hypothesi, relationship, peptid, suggest, opioid, diseas, popul, onset, studi, risk, autist</i>	General terms which summarize what WF-PUB-1998 talks about. Almost all the terms are related to its medical background.
9 (2.6%)	<i>seri, development, parent, report, abnorm, loss, child, autist, associ, acquir, consecut, articl, spectrum, coliti, pervas, symptom, specif, attent, author, count, skill, normal, public, concurr, characterist, claim, abdomin, reduc, enabl, eight</i>	A large number of terms from the medical domain. The connection with WF-PUB-1998 is less evident.
10 (3.9%)	<i>mucos, regress, trigger, subtl, food, medic, symptom, development, pattern, extens, lead, clear, result, retract, disord, specif, condit, enterocol, intoler, possibl, research, also, develop, affect, suggest, case, diarrhea, includ, year, caus</i>	General terms which summarize WF-PUB-1998 medical background. Might also mention its retract

11 (3.7%)	<i>understand, paper, period, peer, cohort, development, singl, review, widespread, technic, major, help, studi, rat, scienc, public, littl, interact, relationship, origin, week, outbreak, neurochem, normal, media, articl, sometim, regress, debat, report</i>	Focuses on technical aspects and doesn't include any medical terminology.
12 (10.8%)	<i>find, retract, uptak, studi, subsequ, media, vaccin, paper, fear, link, controversi, increas, evid, measl, scientif, articl, publish, number, safeti, mump, journal, belief, parent, public, mani, health, mother, concern, claim, lancet</i>	Talks and discusses WF-PUB-1998 retraction phenomena from different perspectives, e.g. social impact. It can also talk about the negative impacts.
13 (3.6%)	<i>lancet, cost, articl, rhetor, health, paper, public, text, scienc, begin, outbreak, emerg, autist, immedi, also, featur, interpret, acquir, origin, caus, controversi, distress, depart, note, might, languag, debat, measl, behavior, vaccin</i>	Talks and discusses WF-RET-CASE from different perspectives, yet far from the medical domain. It might also not take in consideration the paper retraction in the discussion.
14 (5.6%)	<i>unit, publish, studi, state, bowel, possibl, general, paper, measl, immunis, royal, case, press, diseas, report, free, group, three, kingdom, research, link, controversi, hospit, risk, would, receiv, vaccin, women, journal, earlier</i>	Discusses the medical conclusions arised from WF-PUB-1998
15 (3.3%)	<i>intestin, report, associ, studi, regress, retract, team, hospit, bowel, sever, find, ileum, royal, behavior, hypothesi, free, subsequ, caus, ethic, altmetr, problem, colleagu, vaccin, consider, research, connect, abnorm, paper, number, british</i>	Discusses the medical conclusions arised from WF-PUB-1998. It might point out the emerging controversies of the paper.
16 (4.8%)	<i>associ, development, regress, whether, initi, spectrum, vaccin, bowel, trigger, possibl, propos, widespread, autoimmun, autist, disord, specif, environment, diseas, increas, coverag, concern, public, articl, hypothesi, virus, aris, question, base, andrew, sinc</i>	Talks about the medical thematics and conclusions of WF-PUB-1998.
17 (5.4%)	<i>refer, articl, link, vaccin, skill, normal, histori, mump, acquir, side, bowel, support, concern, studi, public, research, suggest, associ, measl, demonstr, paper, lancet, evid, pediatr, symptom, exist, follow, prove, describ, diseas</i>	Discusses the conclusions and impact of WF-PUB-1998
18 (4.6%)	<i>receiv, countri, author, health, measl, articl, year, effect, andrew, eight, recent, parent, behavior, case, suggest, vaccin, paper, begin, safeti, british, sinc, studi, develop, short, mump, patient, campaign, publish, report, investig</i>	Gives an overview of WF-RET-CASE, without necessarily analysing the contents and conclusions.
19 (3.2%)	<i>articl, factor, constip, nonspecif, britain, caus, first, autist, measl, compon, year, upon, call, intestin, publish, dquo, retract, make, immun, event, disord, neurolog, unnessessari, permeabl, find, apoptosi, vaccin, describ, scientif, syndrom</i>	Talks about WF-RET-CASE and the medical conclusions of the paper.

20 (1.6%)	<i>theori, council, nationwid, queri, consequ, profession, medic, continu, ultim, despit, accept, long, case, optim, rat, vitamin, pertussi, myelogenesi, impair, persist, altern, indic, https, prsa, cross, field, coverag, notif, dismiss, collaps</i>	Discusses WF-PUB-1998 as a case of study that might be of interest to a better understanding of the research or more specifically the medical research.
21 (5%)	<i>vaccin, coverag, public, associ, link, articl, suggest, publish, research, media, time, citat, claim, thimeros, across, prove, particular, measl, lancet, lead, whether, parent, paper, evid, subsequ, extens, sinc, mump, around, increas</i>	Talks about WF-RET-CASE and consequences. Might also discuss the paper links and citations.
22 (4.4%)	<i>connect, retract, paper, potenti, topic, disord, research, deer, receiv, exempl, studi, lancet, report, claim, elliman, causal, use, development, inflammatori, although, concern, data, type, exist, inform, bowel, publish, base, measl, attent</i>	Discusses WF-PUB-1998 and retraction from non-medical aspects. Might refer to it as a retraction example.

Table 6. The 22 topics generated from the in-text citations context topic model. For each topic (row) we mention its proportion percentage in the corpus (column 1), and the 30 most relevant terms (column 2), and we give our interpretation of it (column 3).