Check for updates

A qualitative and quantitative analysis of open citations to retracted articles: the Wakefield 1998 et al.'s case

Ivan Heibi[1,2] · Silvio Peroni[1,2]

Abstract

In this article, we show the results of a quantitative and qualitative analysis of open citations 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 The main purpose of our study is to understand the behavior of the publications citing one retracted article and the characteristics of the citations the retracted article accumulated over time.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 visualized the results. The data and services used are all open and free to foster the reproducibility of the analysis. The outcomes concerned the analysis of the entities citing Wakefield et al.'s article 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 either discussing or dealing with the retraction of Wakefield et al.'s article even before its full retraction happened in 2010. Articles in the social sciences domain citing the Wakefield et al.'s one were among those that have mostly discussed its retraction. In addition, when observing the in-text citations, we noticed that a large number of the citations received by Wakefield et al.'s article has focused on general discussions without recalling strictly medical details, especially after the full retraction. Medical studies did not hesitate in acknowledging the retraction of the Wakefield et al.'s article and often provided strong negative statements on it.

✉ Ivan Heibi
  ivan.heibi2@unibo.it

  Silvio Peroni
  silvio.peroni@unibo.it

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

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

Introduction

A peer-reviewed retracted article should be considered as an invalid source of knowledge depending on specific reasons for its retraction which might include scientific misconduct, fabrication, general content errors, plagiarism and self-plagiarism (Moylan et al. 2016). The editor(s) of the venue in which the original publication was published has the final decision about whether or not to retract it. This decision is accompanied by a retraction notice. Also, sometimes a 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 and implemented in the past years to show updated notifications regarding articles, such as retractions and error corrections (Meyer 2011). An important service which keeps track of and collects retractions of scholarly articles is Retraction Watch (http://retractionwatch.com/) (Collier 2011). In addition, the COPE retraction guidelines (Barbour et al. 2009) state that "the original article should not be completely removed or 'replaced', but should be retained and linked to", practically enabling studies on the retracted articles.

The retraction phenomenon has been largely discussed by scientometricians. We can organize the studies in this domain into two macro categories: (a) large scale analysis and (b) case 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, by considering the reasons for retraction introduced in (Bar-Ilan et al. 2018) and (Lu et al. 2013), authors 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 shown also by other works such as (Azoulay et al. 2017) (Mongeon et al. 2016) (Shuai et al. 2017). Along the same lines, Feng et al. (2020) introduced a multi-dimensional observation framework using four dimensions, including scientific impact, technological impact, funding impact and Altmetric impact. Others have discussed possible approaches to avoid retraction and related issues in citing retracted papers. For example, Mott et al. (2019) have suggested strategies to adopt for improving the effectiveness of retraction notices and the awareness of the citing authors. Another example is the work done by Bordignon (2020) which investigated the different impacts that negative citations in articles and comments posted on post-publication peer review platforms have on to the correction of science. Finally, Bolland et al. (2021) recently analyzed citations made before retraction.

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 corpus starting from the findings and results obtained – e.g. by focusing on post-retraction citations and the related sentiment when citing (Bar-Ilan et al. 2017), by classifying citation contexts (Jan et al. 2018) and by running a network analysis study (Chen et al. 2014). Similarly, the work done by Teixeira da Silva et al. (2017) focused on a restricted list of the top ten cited retracted articles for analyzing 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 in order to perform a network analysis which monitors the propagation of the results and findings of the retracted article (van der Vet et al. 2016), to notice whether retracted works are still being cited without mentioning their retraction (Bornemann-Cimenti et al. 2016) and to classify reasons for citing the retracted articles (Luwel et al. 2019). The work done by Schneider et al. (2020) is another example of a work falling into this category of studies.

The analysis we present in this article is close to those introduced in the latter set of studies. 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. 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 analyzed considering its popularity among several anti-vaccine movements and the implications it has had for society (Chen et al. 2013). Since WF-PUB-1998 is also one of the most cited retraction cases, this large quantity of citations will help us have a better assessment of the methodology we introduce in this work.

In our study, we focused on the citation analysis of the WF-PUB-1998 from a quantitative and qualitative point of view. We split the analysis and the findings into three different periods – i.e., P1, P2 and P3, based on the years of the partial (2004) and final (2010) retraction of WF-PUB-1998. In particular P1 refers to the period from WF-PUB-1998 publication to the partial retraction (1998–2004), P2 from the year after the partial retraction to the full retraction (2005–2010) and P3 from the year after the final retraction to 2017 (2011–2017). We considered 2017 as last year due to the availability of the citation data in OpenCitations (Peroni et al. 2020) we gathered at the time this work was performed, i.e., the November 2018 COCI release (OpenCitations 2018). Our workflow follows a precise methodology specifically designed for the application of a qualitative and quantitative analysis on retracted articles. The methodology goes through three main stages: annotation, data and results visualization and answering the research questions.

A similar work on the retraction case of WF-PUB-1998 has been recently also presented by Suelzer et al. (2019). The analysis of Suelzer et al. is based on a collection of 1153 citing works retrieved using Web of Science and the findings are also compliant with the three periods P1-P3 we have used in our work. The collected citations have been classified into negative/ perfunctory/affirmative and annotated as those that have/have not documented the partial/full retraction of WF-PUB-1998. Suelzer et al. suggest that improvements are needed from publishers, bibliographic databases and citation management software to ensure that retracted articles are accurately documented.

Unlike (Suelzer et al. 2019), our work 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. We present a detailed comparison that highlights the method, annotated features and findings of our work and (Suelzer et al. 2019) in the final part of this article.

The aim of our work is to answer to the following research questions:

RQ1   What are the research topics introduced in the articles citing WF-PUB-1998 before and after its retraction?

RQ2   What are the most relevant characteristics of the in-text citations (e.g. intent, sentiment, mention of the retraction, etc.) 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 (subsections "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.

Before describing the steps of our methodology, though, we give a preliminary and brief introduction to open citations, since they represent our main source to gather initial citation data. In particular, we provide their definition and usage, since they represent an important part of our methodology.

Open citations

Following the definition provided in (Peroni et al. 2018a), a bibliographic citation is an open citation when the data needed to define the citation are:

- Structured: expressed in one or more machine-readable format such as JSON.
- Separate: available without the need to access the source article in which the citation is defined.
- Open: freely accessible and reusable without restrictions.
- Identifiable: the entities linked by an open citation must be clearly identified by using a specific persistent identifier scheme, such as a DOI, or a URL.
- Available: it must be possible to obtain the basic metadata of the entities involved in the citation by resolving their identifiers.

The open citation data used in this work are provided by OpenCitations (Peroni et al. 2020), an independent infrastructure organization for open scholarship, which is dedicated to the publication of open citation data using Semantic Web technologies (Berners-Lee et al. 2001), which facilitate the use of precise semantics for the encoding and creation of machine-processable data on the Web.

The Semantic Web technologies used by OpenCitations permit the publication of bibliographic and citation data as Linked Open Data (LOD) (Bizer et al. 2011). These bibliographic and citation data are compliant with the OpenCitations Data Model (Daquino et al. 2020), which is implemented by means of the SPAR Ontologies (http://www.sparontologies.net) (Peroni et al. 2018b). In particular, citations are described using Citation Typing Ontology (CiTO, http://purl.org/spar/cito) (Peroni et al. 2012), which allows one to create metadata describing citations (that are distinct from the metadata describing the cited works themselves) and permits the *intent* of an author when referring to another document to be captured.

In particular, the OpenCitations collection we used to gather all the open citation data is COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations (http://opencitati ons.net/index/coci) (Heibi et al. 2019b), which contains details of all the citations that are specified by the open references to DOI-identified works present in Crossref (Hendricks et al. 2020).

COCI and CiTO are two of the main components in the methodology we present. We use COCI to gather the citations of the retracted article in consideration and we adopt the CiTO definitions to characterize the citation intents, based on the citation context and following a guiding schema.

Data gathering

The data sources we used to gather the data for our analysis were OpenCitations COCI, that we used to retrieve citation data, the RetractionWatch database (http://retractiondatab ase.org/) used to retrieve information of retracted articles, SCImago (https://www.scima gojr.com/) to retrieve subject areas and subject categories of articles and the ISBNDB service (https://isbndb.com/) to look up the Library of Congress Classification code (LCC, https://www.loc.gov/catdir/cpso/lcco/) of books.

We queried the COCI REST API (http://opencitations.net/index/coci/api/v1) when COCI was populated with citation data from 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 a 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 organized in five steps, introduced in Table 1. Additional information and details about the step related to the gathering of the data used in our analysis can be found in (Heibi et al. 2020a), which goes deeper into its technical aspects (e.g. execution of software code and additional contextual information) and does not discuss any aspect related with the other steps of the methodology. Thus, the methodology described herein is self-contained and enables the reproducibility of our work.

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 API (http://opencitati ons.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 to manually check if each of the citing entities (identified by its DOI) has been retracted as well 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. SCImago groups the journals into subject areas (27 major thematic areas, e.g. *Medicine*) and subject categories (313 specific subject categories, e.g. *Immunology and Allergy*). Some venues can have more than one subject area and subject category – we considered all of them in these cases. For publication venues with ISBN (mainly books), we used the ISBNDB service to look up the related Library of Congress

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

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) DOI (1.2) year of publication (1.3) title (1.4) venue id (ISSN/ISBN) (1.5) venue title
(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) is / is not retracted
(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) subject area (3.2) subject category
(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) abstract (4.2) in-text citation section (4.3) in-text citation context (4.4) in-text reference pointer
(5) Annotating the in-text citations characteristics	Manually annotating the intent (based on the citation functions of CiTO) and sentiment of each in-text citation and specifying whether the text in citation contexts mentions the retraction of the cited article	In-text citation contexts	For each in-text citation: (5.1) citation intent (5.2) citation sentiment (5.3) retraction is / is not mentioned

Classification code. Then, we mapped the LCC categories we found to SCImago subject areas and categories 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, which is available at https://github.com/ivanhb/cits-ret-method. If this is true, we automatically annotated the subject area with such value, while we assigned, as subject category, 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 this is the case, 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 needed to be manually annotated by consulting the complete LCC index (http://www.loc.gov/catdir/cpso/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 pointers denoting a bibliographic reference referencing WF-PUB-1998 (e.g. "Wakefield et al. 1998"), the citation contexts of the 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. As suggested in (Ritchie et al. 2008), this strategy for the definition of the citation context window seemed appropriate to guarantee an accurate annotation of the intents of citations. Special cases/exceptions to this rule (e.g. if the sentence that contains the in-text reference is the first sentence of a section, then the in-text citation context did not include the preceding sentence) are treated in (Heibi et al. 2020a). Also, we characterized each of the sections containing in-text citations according to their 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 analyzed each citation context of the in-text citation retrieved and we inferred:

- the perceived sentiment regarding WF-PUB-1998;

- whether at least one citation context of any in-text citation of the citing entity explicitly mentions 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. the citing entity uses a method defined in the cited entity).

For specifying the citation sentiment, we followed the classification proposed by (Bar-Ilan et al. 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 (Peroni et al. 2012). 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. In our methodology, we made a clear distinction between the *sentiment* and the *intent* of the citation, since the annotation of a specific citation intent does not directly imply its sentiment. For instance, the intent might be *to obtain background from* the cited entity, yet this could be done with a negative/positive perception toward it. For instance, the authors of the following in-text citation cited WF-PUB-1998 to obtain background information from it and they expressed a slightly negative sentiment toward it: "We explain one example of single-source overlays in detail here. The seed

Fig. 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. The first large row contains the three macro-categories: (1) "Reviewing …", (2) "Affecting …" and (3) "Referring …". Each macro-category has at least two subcategories and each subcategory refers to a set of citation functions. The first row defines what are the citation functions suitable for it through the help of a guiding sentence which needs to be completed according to the chosen sub-category and citation function

article in the example, Wakefield et al. (1998), is a highly cited retracted article, which has profound implications on public health, especially on vaccine uptakes from children" (Gap Analytics 2014).

We performed a manual annotation of the in-text citations using the decision model we developed for this study, summarized in Fig. 1. The decision model is organized into three main macro categories (i.e., large columns): (1) "Reviewing …", (2) "Affecting …" and (3) "Referring …". Each macro category has at least two other inner classifications (i.e., columns). For instance, the "Affecting" macro-category has the "citing entity" and "cited entity" inner columns. The macro categories and their inner classifications work as guiding schemas for the annotator and are not part of the final annotations.

The 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 Fig. 1);
2. for each macro-category selected, we selected one or more citation functions choosing between those provided by CiTO – the decision model provides a template and an example (i.e.,*"Fill the sentence …"*) to help us chose the most suitable one;
3. if we selected only one citation function, we annotated the in-text citation intent with that 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 x-axis (from 1 to 8) and in the y-axis (from 10 to 50), as shown in Fig. 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.

Topic modeling

Some recent works such as (Bornmann et al. 2020) and (Crothers et al. 2020) analyzed the context of citations to highly cited articles that have explained and introduced important concepts. The idea of these works is to count the number of times the concepts are mentioned in the citation context. This analysis only makes sense if the highly cited publications have introduced at least one important concept. The work done by Lyu et al. (2020) examined the Big Data research domain and investigated how the academic topics shift across altmetric sources (e.g. Twitter). Another recent work proposed by Zhang et al. (2021) analyzed the topic evolution in early COVID-19 research. Generally, an analysis toward the topic evolution in specific research domains, institutions, periods, or following an important event (e.g., the COVID-19 pandemic), is an important subject of interest for the development of effective ways to inform research strategies and evaluate research activities, as it is demonstrated by the development of tools such as Elsevier's SciVal (https://www.elsevier.com/solutions/scival), a web-based tool for visualizing and investigating these aspects.

In our study, we wanted to generalize to go beyond a set of popular concepts and we tried to consider an arbitrary number of concepts/topics that we want to identify using a computational approach. We decided to address this problem by using a topic modeling technique, which is an appropriate method to use for the automatic analysis of texts that

works without having preliminary knowledge on the subjects the texts are about. For instance, a recent application of topic modelling for similar purposes is described in the work by Han (2020), who used it to investigate the evolution of research topics in the library and information science (LIS) domain.

We run a topic modeling analysis on the textual features we gathered (i.e. abstracts and citation contexts) using MITAO (https://github.com/catarsi/mitao) (Ferri et al. 2020), a visual interface to create a customizable visual workflow based on the Latent Dirichlet Allocation (LDA) topic modeling (Jelodar et al. 2019). In particular, the topic modeling analysis we introduce in this article resulted in the creation of two topic models, one for the abstracts and one for the in-text citation contexts. This analysis is based on three main stages: (a) the identification of the number of topics to consider given a corpus of texts, (b) building the topic modeling workflow and (c) generating the results and the related visualizations. We discuss each of these stages individually in the following subsections.

Number of topics

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 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.

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 citation contexts of the entities that have cited WP-PUB-1998. The coherence score plateaued around 22–23 topics. Thus, we decided to consider 22 topics for the citation contexts. We have used a similar approach for abstracts, in this case the coherence score plateaued around 13–14 topics.
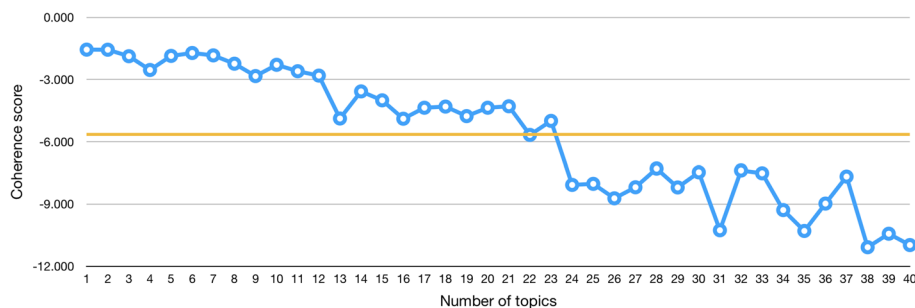

Fig. 2 The coherence score of different LDA topic models built using a variable number of topics, from 1 to 40. The topic model is based on the corpus and dictionary of the in-text citation contexts. The orange line is the average value and it plateaus around 22–23 topics

The topic modeling workflow

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 punctuation, unnecessary characters and stopwords. In our study, 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 lemmatize 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 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 citation contexts and we gave as input the number of topics, identified using the coherence score, to each model following the results of the previous stage. Figure 3 illustrates graphically the workflow we developed and run using MITAO. Some of the components of the workflow are used to generate the results and the visualizations, which are introduced in the following subsection.
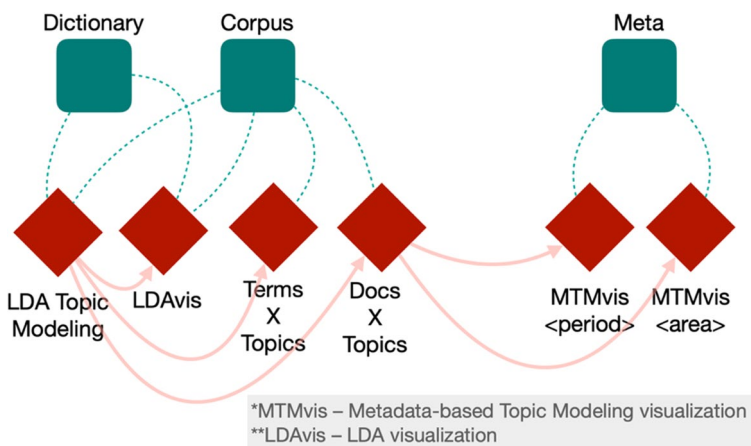

Fig. 3 The workflow, created via MITAO, we used for computing the LDA topic modeling and generating the LDAvis (LDA visualization) and MTMvis (Metadata-based Topic Modeling visualization) visualizations (the tools "LDAvis", "MTMvis < period >" and "MTMvis < area >"). The green squares are used to specify input material which is considered by the various tools composing the workflow (i.e., the red rhombi). In particular, the workflow takes three inputs: (a) the vectorized corpus ("Corpus"), (b) a dictionary of words based on the tokenization results ("Dictionary") and (c) the metadata of the original documents forming the corpus ("Meta"). The arrows between the tools indicate the direction of the data flow and the output-input relation among them. For instance, the execution of the workflow starts with the tool "LDA Topic Modeling", that takes in input the "Corpus" and the "Dictionary" and produces an output that is used as part of the input for other three tools, i.e. "LDAvis", "Terms X Topics" and "Docs X Topics"

Results and visualizations

As anticipated in the previous subsections, we used MITAO to generate two datasets for each case (abstracts and citation contexts). Each dataset contained:

- 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;
- document 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 et al. 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 et al. 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 customizable metadata set specified for those documents. We created two MTMvis visualizations for both the abstracts and the citation contexts of the entities citing WF-PUB-1998. The first one shows the representativeness of the topics based on the year of publication while the second takes into consideration the subject area value.

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 analyze 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 analyze 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 behavior 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 analyze these features against the outcomes of the topic model of the in-text citation contexts.

Results

In this section, we present the results of our analysis. All the data and visualizations are available in (Heibi et al. 2020b). Although in this article we present a screenshot of the visualizations, these are provided in dynamic HTML documents and each visualization can be customized using the filters and parameters it makes available. We provide 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 organize 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.

Figure 4 introduces some descriptive statistics of the values described in Table 2. The charts are organized 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.

The second column of Fig. 4 shows a continuous increment in the number of citations and a higher percentage of entities mentioning the retraction. In 2009, we had the smallest percentage of entities which have mentioned the retraction (7%), while we observed the higher percentage value in 2017 (61%). Considering the distribution of the areas of study, we observed a slightly decreasing presence of the *medicine* area in favor of other areas of study which gained much more relevancy in P2 and P3 (e.g. *social sciences* being 1.61%, 8.93% and 12.59% in P1, P2 and P3 respectively). In addition, we noticed the emerging of new areas in P2 and P3, such as *economics* and *environmental science*.

As anticipated in the previous section, we obtained the topic model using the abstracts of all the publications considered and summarized in Fig. 2. Considering the results of the coherence score introduced in Section "Topic Modeling", we built a topic model of 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

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

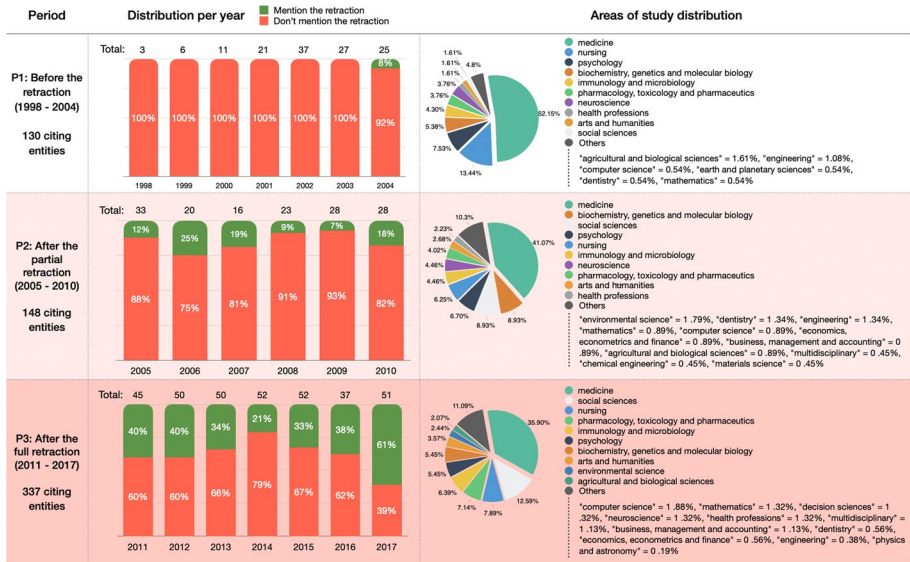| WF-PUB-1998 citing entities features	Values
doi The DOI of the citing article	Total: All the citing entities had a value specified
year The year of publication of the citing article	Total: All the citing entities had a value specified Values: From 1998 (year of publication of WF-PUB-1998) to 2017
title 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 pharmaceutics" (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)

Fig. 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. (Color figure online)
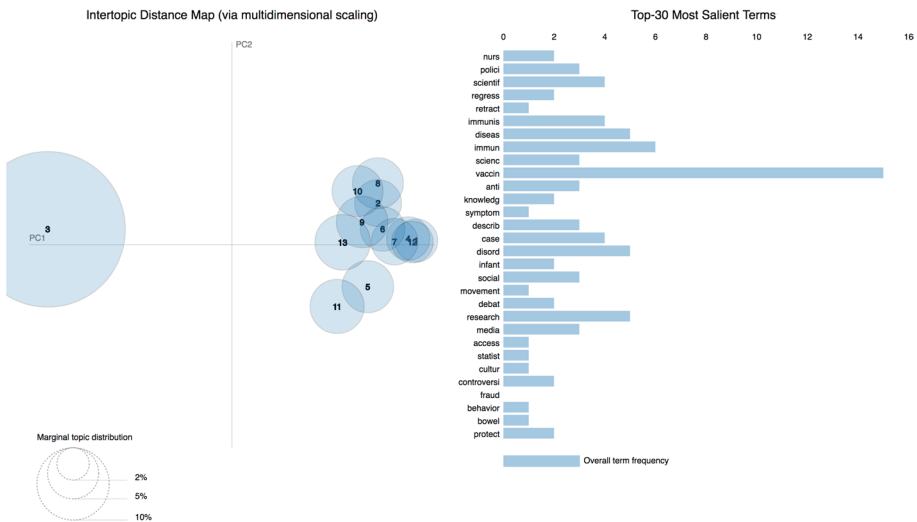

Fig. 5 The LDAvis visualization built over the topic model obtained from the abstracts of the citing entities
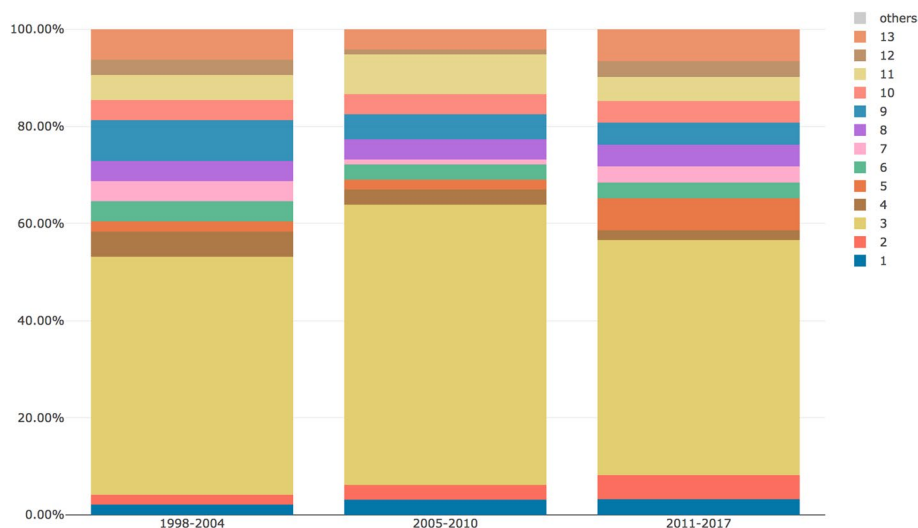
Fig. 6 MTMvis built on the topic model obtained from the abstracts of the citing entities, shown against the three period P1-P3. For each period the visualization plots the topics distribution (e.g. topic 3 is the dominant topic in all the periods: P1, P2 and P3)
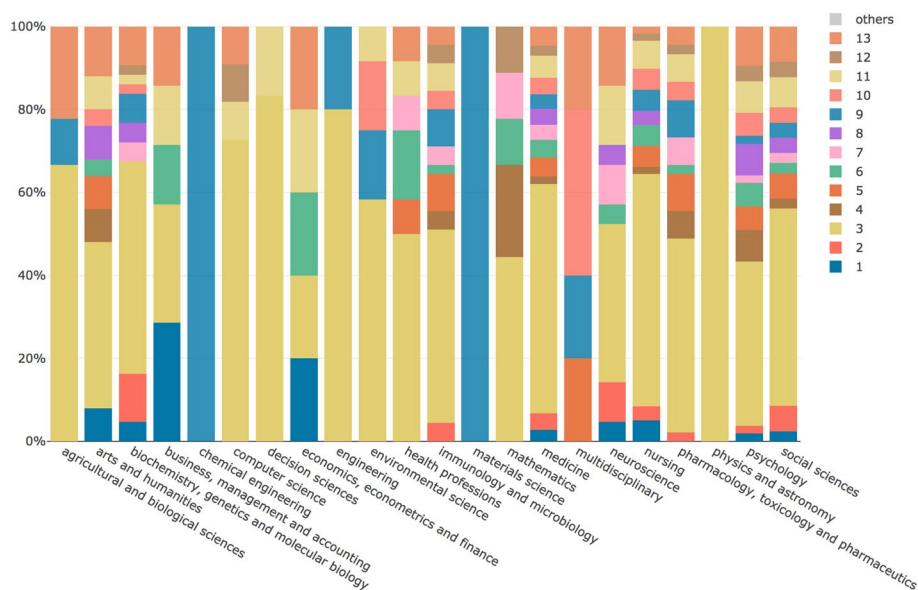

Fig. 7 MTMvis built on the topic model obtained from the abstracts of the citing entities, shown against their subject areas. For each subject area the visualization plots the topics distribution (e.g. topic 3 is the dominant topic in"arts and humanities")

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 phenomenon. 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".

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

The MTMvis visualizations are plotted considering the period P1–P3 (Fig. 6) and the subject areas of the citing articles (Fig. 7). As shown in Fig. 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 Fig. 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.

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

WF-PUB-1998 in-text citations features	Values
intext_citation.section 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: *introduction* (166), *discussion* (61), *results* (28), *background* (36), *conclusions* (17), *method* (15), *abstract* (5)
intext_citation.context The textual context in the citing entity which includes the in-text citation	Total: all the in-text citations had a value specified
intext_citation.pointer 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
intext_citation.intent 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: *discusses* (226), *disputes* (114), *credits* (95), *cites for information* (90), *cites as evidence* (74), *qualifies* (70), *describes* (60), *obtains background from* (56), *critiques* (55), *includes excerpt from* (8), *obtains support from* (6), *uses data from* (5), *uses conclusions from* (4), *ridicules* (4), *extends* (1), *updates* (1), *refutes* (1)
intext_citation.sentiment 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: *neutral* (549), *negative* (300), *positive* (21)

Fig. 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 citation sections distribution

In-text citations

The total number of in-text citations to WF-PUB-1998 gathered from the 615 citing entities was 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.

Figure 8 shows descriptive statistics of some of the values introduced in Table 3. The sentiment is combined with all the statistics displayed (red for negative, yellow for neutral, green for positive). The first column contains the three periods P1–P3 considered in our analysis. The second column shows the distribution per year of the in-text citations, the third column shows the distribution of citation intents and the fourth column shows the distribution of the sections where in-text citations were contained. The sections are classified considering the list proposed in (Suppe 1998) when possible, while all the others are grouped under the label *"Others"* (i.e. a section with a generic title which could not be identified in any section from the proposed list in (Suppe 1998)).

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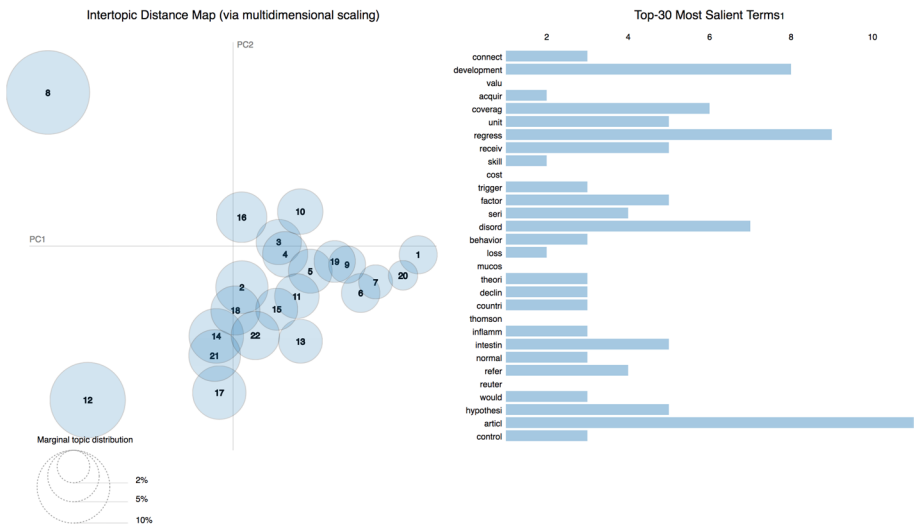
Fig. 9 The LDAvis visualization of the topic model created using the citation contexts of the in-text citations contained in the entities citing *WF-PUB-1998*
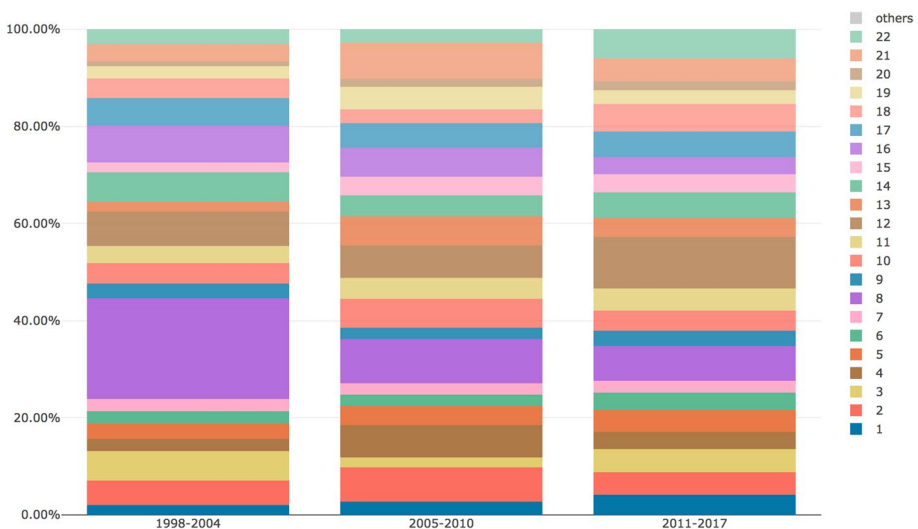

Fig. 10 MTMvis created considering the topics extracted from the citation contexts of the in-text citations citing WF-PUB-1998 according to the periods P1-P3. For each period the visualization plots the topics distribution – e.g., topic 8 (in purple) is the dominant topic in P1

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.

The MTMvis visualizations in Figs. 10 and 11 refer again to the distribution of the topics over P1-P3 and considering the subject areas of the entities containing the in-text

Fig. 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. For each period the visualization plots the topics distribution – e.g., topic 3 (in dark yellow) is the dominant topic of the "arts and humanities" subject area

citations analyzed. 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 behavior 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. In addition, 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*.

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

In this section we address RQ1: what are the research topics introduced in the articles citing WF-PUB-1998 before and after its retraction?

From a quantitative point of view, while looking at the subject areas of the citing entities we gathered (see Fig. 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. 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 (i.e. *medicine*).

In addition, we also noticed a continuous increase in the percentage of entities that have explicitly mentioned the retraction of WF-PUB-1998 in their citation contexts over the time (see Fig. 4). The largest number of citations mentioning the retraction (61%) was in 2017 (the last year we have considered). A considerable percentage of entities mentioned the retraction even before the full retraction notice (e.g. 25% of entities in 2006). So both the full and partial retractions were acknowledged by the citing entities. Indeed some acknowledged the retraction after the partial retraction and before the full retraction notice was raised. This aspect might be also related to the 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 themes 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 medical jargon.
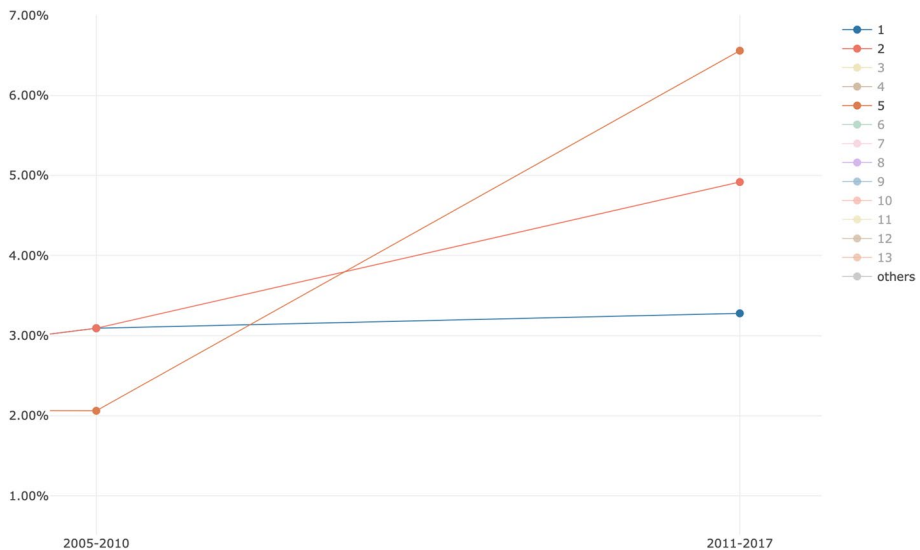

Fig. 12 The evolution of topics 1, 2 and 5 during P2–P3 on all the subject areas plotted using MTMvis. MTMvis has been generated from the topic model created using the abstracts of the citing entities. The themes covered by these topics are close to the retraction phenomena and used a limited number of terms from medical jargon
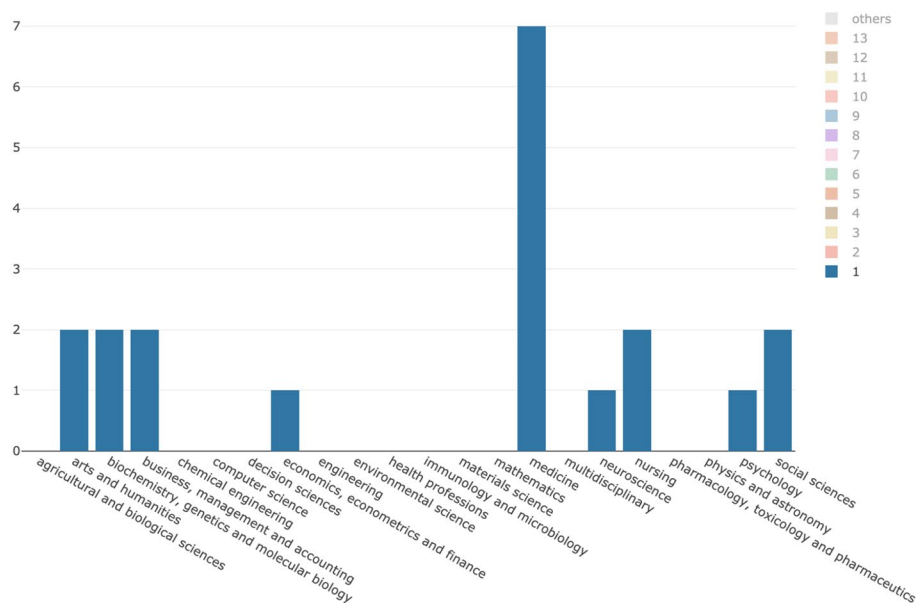
Fig. 13 The distribution of topic 1 over all the subject areas during P2–P3 plotted using MTMvis. MTMvis has been generated from the topic model created using the abstracts of the citing entities. Topic 1 include terms from the social science domain and relates to ethical themes
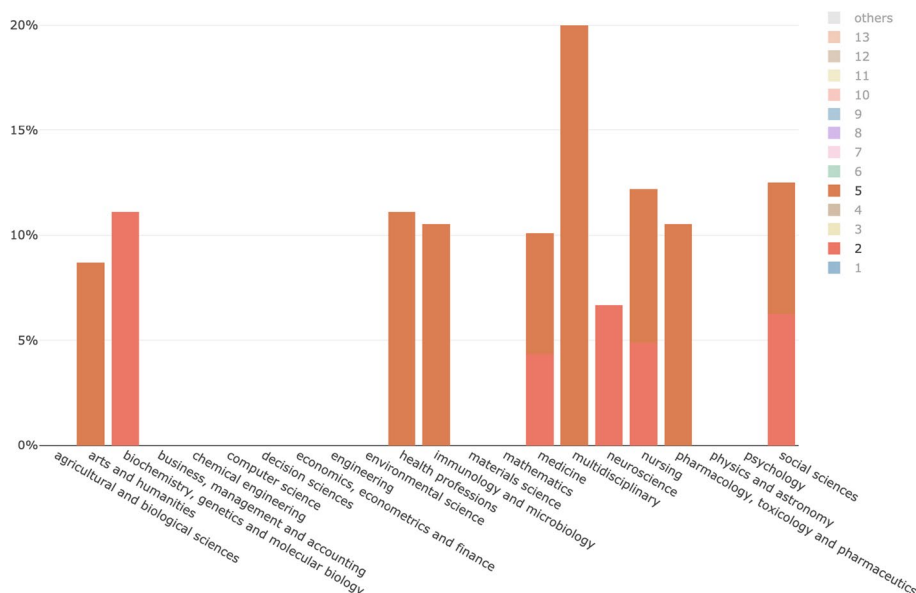

Fig. 14 The subject areas of citing entities published in P2–P3 which includes either topic 2, or 5 in their top 5 topics. The themes covered by these topics relate to the retraction phenomena and use a limited number of terms from medical jargon

A deeper investigation of the evolution of topics 1, 2 and 5 during P2–P3 on all the subject areas, showed that topics 2 and 5 increased significantly in P3 (11.48% vs 5.15%) while topic 1 has a slighter increment (3.09% vs. 3.28%), as we can see in Fig. 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 cited 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 interesting: indeed, topic 1 has little engagement with the medical themes, 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.

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 summarized in Fig. 14. We can see that topics 2 and 5 were listed in the top five topics of twelve subject areas. Avoiding considering 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* subject area 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 themes 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 analyzed.

Answering RQ2

In this section we address RQ2: what are the most relevant characteristics of the in-text citations (e.g. intent, sentiment, mention of the retraction, etc.) in the articles citing WF-PUB-1998 before and after its retraction?

Figure 8 shows that the intended sentiment carried in the citation contexts of the in-text citations referring to WF-PUB-1998 moved to the negative spectrum over time. However, the retraction of WF-PUB-1998 was not always mentioned in these cases. Indeed, as shown in Fig. 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 generic intents such as *discusses* and *cites for information*. This could be 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 with a generic intent without recalling strictly medical details in their text, which are out of the scope of their research domains.
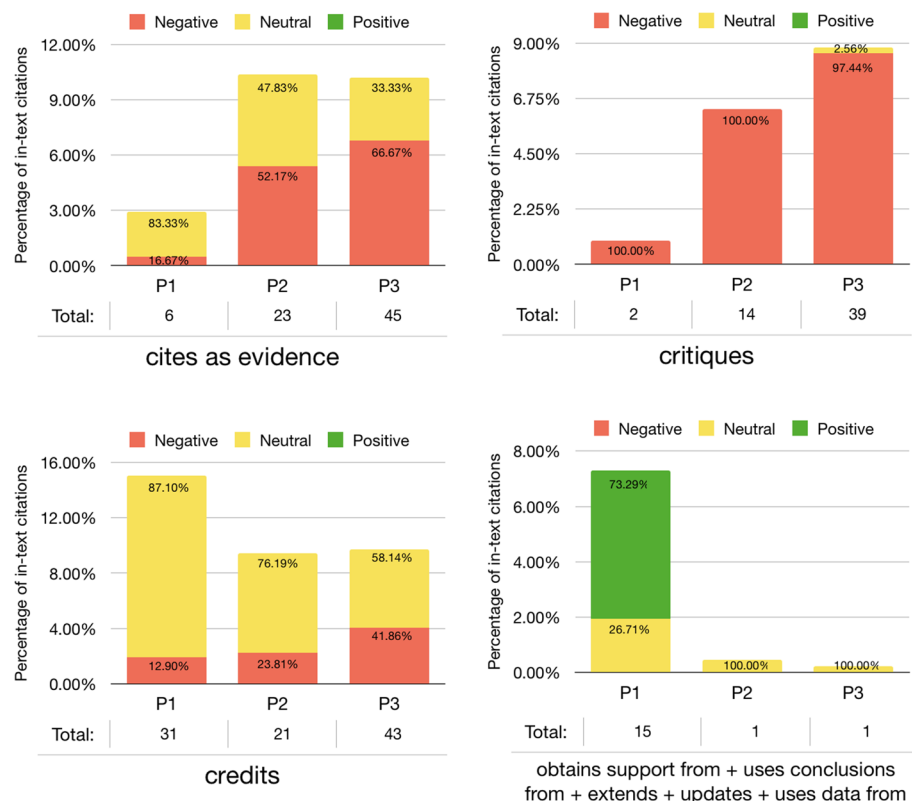
Fig. 15 The four graphs illustrate the way the use of citation intents changed over time (i.e., the three periods P1, P2 and P3) and according to their perceived sentiment. The citation intents *cites as evidence*, *critiques* and *credits* are illustrated in separated charts, that show an increment in the negative sentiment along the three periods

As shown in Fig. 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 and colleagues 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 faulty science.

In Fig. 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.
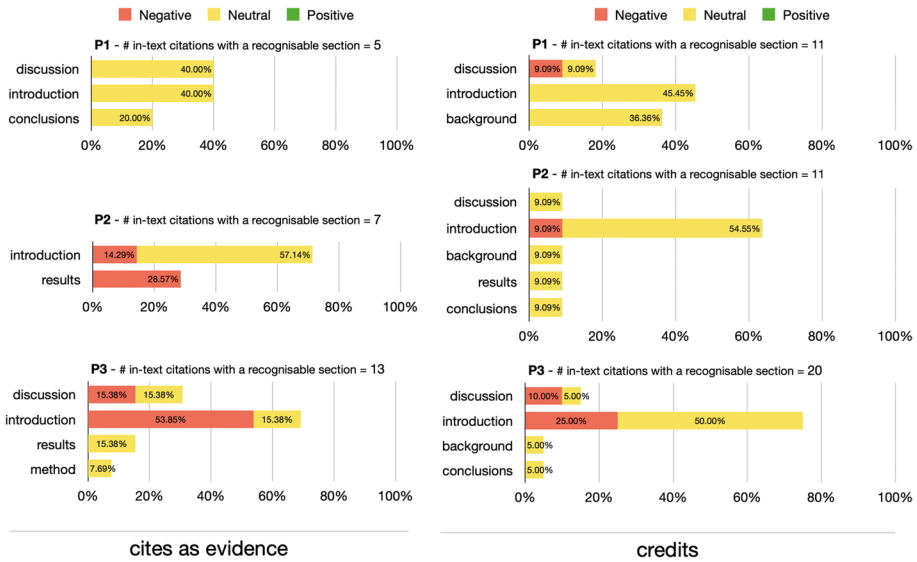
Fig. 16 The *cites as evidence* and *credits* citation intents distributions among the sections (the recognizable ones) and during the three periods (i.e. P1–P3)
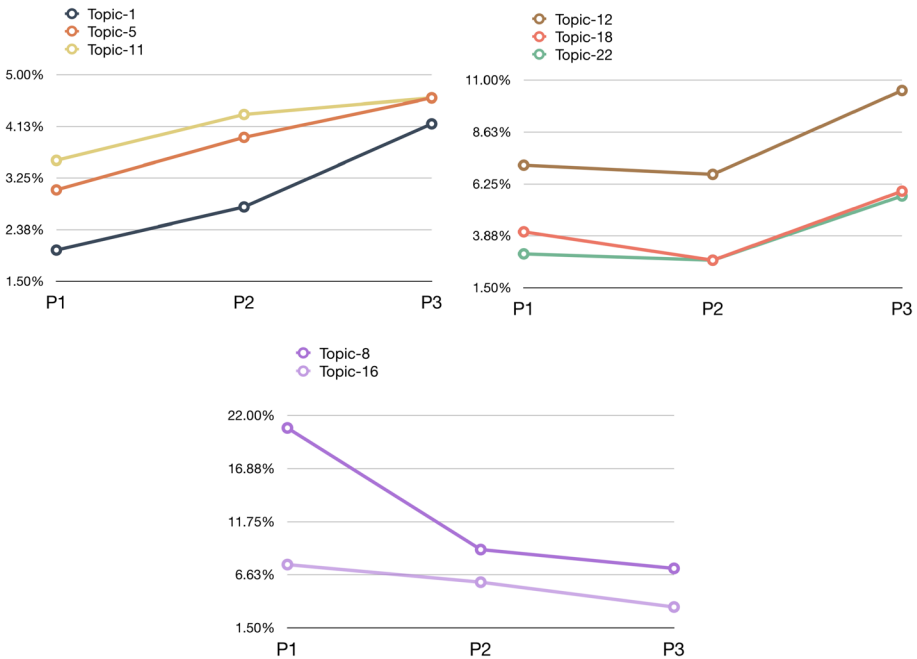

Fig. 17 The evolution over time of three groups of topics defined from the citation contexts of the in-text citations to WF-PUB-1998

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 introduced in Fig. 8: the in-text citations have been concentrated in few sections mostly of descriptive type – i.e. *introduction* and *discussion*.

We analyzed 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 Fig. 17, we focused 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 decrease 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.

Similar behavior 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 themes treated in WF-PUB-1998. The fact that these topics had a clear decrease over time suggests that the most recent citing works provided partial and limited acknowledgement of the conclusions and medical arguments in WF-PUB-1998.

In Fig. 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 Fig. 18 are also included in Fig. 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 Fig. 17, even if it seemed relevant when we focus only on the *medicine* subject area. We had a
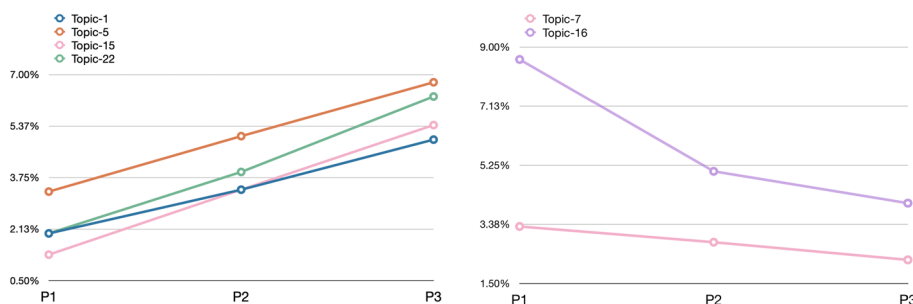

Fig. 18 The increasing (left) and decreasing (right) topics of the in-text citation topic model, considering only the *medicine* area of study

similar situation also with the topics decreasing over time. Indeed, topic 7 (which summarizes WF-PUB-1998 and medical conclusions) is not highlighted in Fig. 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 by pointing out its controversies from a medical perspectives. On the other hand, the decreasing relevance of topic 7 indicates that the entities part of the *medicine* area of study addressed less the medical arguments of WF-PUB-1998 and rather focused on citing and discussing the retraction of WF-PUB-1998 without deepening into its actual content (e.g. method and findings).

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 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 citing entities that do not have 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 from analyzing 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 1211 articles gathered from the Web of Science Core Collection in March 2019, while we collected citations coming from 615 articles in total up till November 2018. This number disparity is strictly related to the prior date, along with the fact that we relied only open citations repositories (COCI in particular) in order to foster the reproducibility of our analysis. We are aware that this might have an impact regarding the final results, although we believe that using openly available resources strengthens the reproducibility of the results and will help us enforce our methodology.

The gap between open and non-open citations should be significantly reduced in the next releases of COCI, due to (a) the recent decision of Elsevier of making reference lists of all articles openly available via Crossref (https://www.elsevier.com/connect/advancing-responsible-research-assessment) and (b) the Crowdsourcing Open Citations Index (CROCI) (Heibi et al. 2019a) which enables scholars and publishers to provide their own citations to OpenCitation to upload them into the OpenCitations Indexes. In principle, these activities will increase the number of citations that can be gathered using our proposed methodology. Other similarities and differences between our study and Suelzer et al.'s one are introduced in Table 4.

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.

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

Feature	Method and results in (Suelzer et al. 2019) compared with our study	
	Method	Results
Analyzed entities and the citation context	Differences: 1) The analysis does not address and collect the in-text citations as separated entities 2) The reviewed in-text citation contexts are not defined following a precise structure/rule – in our work, indeed, the context is defined by the anchor sentence, plus the prior and subsequent sentence Similarities: 1) the citing works are collected and analyzed as individual entities 2) the definition of three periods following the years of the partial and full retraction (i.e. P1, P2 and P3)	Differences: 1) The total number of citing entities is higher – 1,153 (compared to the 615 of our work) 2) The analysis does not mention any statistics regarding the number of in-text citations Similarities: 1) The percentage of citing entities gathered starting from the partial retraction to the last date considered (i.e. March 11, 2019) with respect to the total number is 76% (close to the 78% of our work). Generally, considering the periods P1, P2 and P3 the distribution is respectively 23%, 28% and 49% (close to our distribution: 22%, 24% and 54%)
Mentions of the retraction	Differences: 1) There is no definition of the textual contexts gathered in order to establish whether the citing work had mentioned the retraction or not 2) It also checks whether the retraction is mentioned in the reference list Similarities: 1) A citing entity mentions the retraction, only if the word "retraction" (and its derivatives) is used	Differences: 1) A higher percentage of entities mention the retraction after the partial retraction – i.e. P2-P3 (56% vs 25%). This gap is reduced when considering only the retraction mentioned in the citation context – 38.3% vs 25% Similarities: 1) The overall 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

Table 4 (continued)

Feature	Method and results in (Suelzer et al. 2019) compared with our study	
	Method	Results
Citation Intent	Differences: 1) The citations are characterized into 8 different categories, following the definitions in (Bornmann et al. 2008) 2) There is no distinction between the intent and the sentiment of the citation (e.g. a citation could be characterized as *methodologic* or *negative*, but there is no way to combine both these dimensions) 3) The citation characteristics address the citing works entities (on our work the annotation refer to the specific in-text citations) Similarities: 1) The annotation was performed following a set of rules which guided and helped the annotator	Differences: 1) The plotted results combine in the same dimension both the sentiment and the citation function, following the established annotation rules, thus this fact makes the comparison with our results difficult
Citation sentiment	Differences: 1) It is not part of the analyzed features, although this subject is addressed in the discussion section 2) This information is embedded in the citation intent value (e.g. using the characteristics *negative*, or *affirmative*)	
Citation section	Differences: 1) It is not part of the analyzed features, although this subject is part of the discussion since some articles are provided as examples of citations appearing in the *introduction* and *discussion* sections	
Text analysis (i.e. topic modeling)	Differences: 1) It is not part of the analyzed features. They give a brief summary regarding their observations of some examples that have cited WF-PUB-1998	

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, on the generation of a citation network starting from either our seed retracted article or from its citing entities, as suggested by van der Vet et al. (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 article, we have presented the outcomes of a citation analysis of a highly cited and popular retracted article: WF-PUB-1998 (Wakefield et al. 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 analyze. We approached our general goal through the definition of two research questions aiming at analyzing 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 (using a topic modelling technique) and, thus, to address the research questions.

Our results have been presented according to the entities we analyzed: entities citing WF-PUB-1998 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 introduce 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 we compared our methods and results with the ones in (Suelzer et al. 2019). The bigger difference has regarded the additional features we have considered in our analysis – i.e. the citation sentiment, the citation section and the topic modeling analysis. Many of our findings have also confirmed the results of Suelzer et al.'s work.

Appendix

See Tables 5 and 6.

Table 5 The 13 topics available in the topic model obtained from the abstracts of the articles citing WF-PUB-1998. 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 (3.5%)	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	Close to the social studies domain, might take in consideration ethical thematic
2 (4.2%)	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	Includes terms related to the retraction phenomena (study-domain independent), it includes terms used in scientometric analysis
3 (46%)	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	This is by far the largest 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%)	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	Includes terms from the Medical and pharmaceutical field
5 (5.1%)	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	Some terms are out of the medical field of study and might indicate a possible discussion
6 (3.8%)	statist, cultur, describ, infant, citat, exampl, articl, symptom, case, american, journal, metabol, combin, parent, disagr, doctor, abl, associ, construction, bordetella, basi, advers, illustr, gliadin, illusori, literatur, indic, unit, eat, rather	A high number of terms are related to the scientometric field of study. All the terms are objective and do not indicate an opinion or a discussion
7 (4.1%)	biomark, altmetr, nan, behavior, disord, occur, well, postpon, answer, herd, context, fear, genet, graphic, appeal, interquartil, order, vaccin, gene, sensori, evalu, modul, geneticist, chapter, lymphocyt, abstain, putat, approach, protect, homeopathi	Includes terms from the biology, pharmacology and genetics field of study. Might also indicate a statistical analysis along with an open discussion
8 (4.9%)	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	A large part of the terms are close to WF-PUB-1998 treated thematics. Mostly from the medical field of study

Table 5 (continued)

Topic (proportion)	Terms (the 30 most probable terms)	Interpretation
9 (5.1%)	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	Concern the retraction phenomena, followed with some medical expressions. It also includes strong terms such as "fraud"
10 (4.9%)	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	General terms related to WF-PUB-1998 treated thematics, some are cor-related with a discussion around the retraction phenomena
11 (5.6%)	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	Terms from the medical study field related to WF-PUB-1998 treated themat-ics
12 (3.3%)	uncertainti, 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	Includes terms close to the computer science lexicon and from the pediatric field of study
13 (5.9%)	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	Includes a large number of general terms from different fields of study, part of them are correlated with WF-PUB-1998 thematics

Table 6 The 22 topics available in the topic model generated using the in-text citation contexts contained in the articles citing WF-PUB-1998. 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%)	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	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%)	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	General terms which summarize what WF-PUB-1998 talks about
3 (3.9%)	articl, case, colon, development, lancet, enterocol, report, assert, associ, public, vaccin, disord, follow, signific, diagnosi, base, group, mump-srubella, treatment, controversi, parent, scientif, evid, result, symptom, studi, publish, link, reaction, unknown	General terms which summarize what WF-PUB-1998 talks about
4 (3.9%)	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	General terms which summarize what WF-PUB-1998 talks about. This topic might also outline other information related to the paper
5 (3.6%)	vaccin, link, research, suggest, report, appear, measl, focus, scientif, develop, univers, reject, possibl, studi, associ, mump, investig, though, fund, even, continu, around, wakefieldet, newspap, result, doubt, relat, signific, high, evid	Discusses the controversy around WF-RET-CASE
6 (2.8%)	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	Does not include any medical term, it rather focuses on other related aspects concerned with WF-PUB-1998
7 (2.2%)	diseas, caus, characterist, process, read, inflamm, alarm, report, assert, without, show, ileocolon, bowel, declin, safeti, student, first, exampl, intestin, media, young, detail, autist, adult, possibl, scientif, studi, even, control, wide	General terms which summarize what WF-PUB-1998 talks about. Many terms are related to its medical background
8 (13.2%)	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	General terms which summarize what WF-PUB-1998 talks about. Almost all the terms are related to its medical background

Table 6 (continued)

Topic (proportion)	Terms (the 30 most probable terms)	Interpretation
9 (2.6%)	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	A large number of terms from the medical domain. The connection with WF-PUB-1998 is less evident
10 (3.9%)	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	General terms which summarize WF-PUB-1998 medical background. Might also mention its retract
11 (3.7%)	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	Focuses on technical aspects and does not include any medical terminology
12 (10.8%)	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	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%)	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	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%)	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	Discusses the medical conclusions arised from WF-PUB-1998
15 (3.3%)	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	Discusses the medical conclusions arised from WF-PUB-1998. It might point out the emerging controversies of the paper

Table 6 (continued)

Topic (proportion)	Terms (the 30 most probable terms)	Interpretation
16 (4.8%)	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	Talks about the medical thematics and conclusions of WF-PUB-1998
17 (5.4%)	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	Discusses the conclusions and impact of WF-PUB-1998
18 (4.6%)	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	Gives an overview of WF-RET-CASE, without necessarily analysing the contents and conclusions
19 (3.2%)	articl, factor, constip, nonspecif, britain, caus, first, autist, measl, compon, year, upon, call, intestin, publish, dquo, retract, make, immun, event, disord, neurolog, unnecessari, permeabl, find, apoptosi, vaccin, describ, scientif, syndrom	Talks about WF-RET-CASE and the medical conclusions of the paper
20 (1.6%)	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	Discusses WF-PUB-1998 as a case study that might be of interest to a better understanding of the research or more specifically the medical research
21 (5%)	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	Talks about WF-RET-CASE and consequences. Might also discuss the paper links and citations
22 (4.4%)	connect, retract, paper, potenti, topic, disord, research, deer, receiv, exampl, studi, lancet, report, claim, elliman, causal, use, development, inflammatori, although, concern, data, type, exist, inform, bowel, publish, base, measl, attent	Discusses WF-PUB-1998 and retraction from non-medical aspects. Might refer to it as a retraction example

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

Barbour, V., Kleinert, S., Wager, E., & Yentis, S. (2009). *Guidelines for retracting articles*. Committee on Publication Ethics. https://doi.org/10.24318/cope.2019.1.4

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

Bar-Ilan, J., & Halevi, G. (2018). Temporal characteristics of retracted articles. *Scientometrics, 116*(3), 1771–1783. https://doi.org/10.1007/s11192-018-2802-y

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.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web. Scientific American, 284*(5), 34–43.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts* (pp. 205–227). IGI global. https://doi.org/10.4018/jswis.2009081901

Bolland, M. J., Grey, A., & Avenell, A. (2021). Citation of retracted publications: A challenging problem. *Accountability in Research*. https://doi.org/10.1080/08989621.2021.1886933

Bordignon, F. (2020). Self-correction of science: A comparative study of negative citations and post-publication peer review. *Scientometrics*. https://doi.org/10.1007/s11192-020-03536-z

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

Bornmann, L., Wray, K. B., & Haunschild, R. (2020). Citation concept analysis (CCA): A new form of citation analysis revealing the usefulness of concepts for other researchers illustrated by exemplary case studies including classic books by Thomas S. *Kuhn and Karl r. Popper. Scientometrics, 122*(2), 1051–1074. https://doi.org/10.1007/s11192-019-03326-2

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

Chen, C., Hu, Z., Milbank, J., & Schultz, T. (2013). A visual analytic study of retracted articles in scientific literature. *Journal of the American Society for Information Science and Technology, 64*(2), 234–253. https://doi.org/10.1002/asi.22755

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

Crothers, C., Bornmann, L., & Haunschild, R. (2020). Citation concept analysis (CCA) of Robert K. Merton's book "Social theory and social structure": How often are certain concepts from the book cited in subsequent publications? Quantitative Science Studies, 1–29. https://doi.org/10.1162/qss_a_00029

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

Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., ... & Zumstein, P. (2020). The OpenCitations data model. In *International Semantic Web Conference* (pp. 447–463). Springer, Cham. https://doi.org/10.1007/978-3-030-62466-8_28

Feng L., Yuan J., & Yang L. (2020). An observation framework for retracted publications in multiple dimensions. Scientometrics, 125(2), 1445–1457. Scopus. https://doi.org/10.1007/s11192-020-03702-3

Ferri, P., Heibi, I., Pareschi, L., & Peroni, S. (2020). MITAO: A user friendly and modular software for topic modelling. *PuntOorg International Journal, 5*(2), 135–149. https://doi.org/10.19245/25.05.pij.5.2.3.

Gap Analytics. (2014). In C. Chen, *The Fitness of Information* (pp. 217–251). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118858080.ch7

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics, 125*(3), 2561–2595. https://doi.org/10.1007/s11192-020-03721-0

Heibi, I., & Peroni, S. (2020a). 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. (2020b). Methodology data of "A qualitative and quantitative citation analysis toward retracted articles: A case of study." Zenodo. https://doi.org/10.5281/ZENODO.4323221

Heibi, I., Peroni, S., & Shotton, D. (2019a). Crowdsourcing open citations with CROCI-An analysis of the current status of open citations and a proposal. In proceedings of the 17th International conference on Scientometrics & Informatics (ISSI2019)

Heibi, I., Peroni, S., & Shotton, D. (2019b). 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

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies, 1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Jan, R., Bano, S., Ikhlaq ur Rehman, S., & Mehraj, M. (2018). Context analysis of top seven retracted articles: Should Retraction Watch revisit the list? Library Philosophy and Practice, 1–20. 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

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

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

Lyu, X., & Costas, R. (2020). How do academic topics shift across altmetric sources? A case study of the research area of Big Data. *Scientometrics, 123*(2), 909–943. https://doi.org/10.1007/s11192-020-03415-7

May, C., Cotterell, R., & Van Durme, B. (2019). An Analysis of Lemmatization on Topic Models of Morphologically Rich Language.

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., & Kowalczuk, M. K. (2016). Why articles are retracted: A retrospective cross-sectional study of retraction notices at BioMed Central. *British Medical Journal 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) . 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

Peroni, S., & Shotton, D. (2018a). Open Citation: Definition. Figshare. Journal Contribution. https://doi.org/10.6084/m9.figshare.6683855.v1

Peroni, S., & Shotton, D. (2018b). The SPAR ontologies. In *International Semantic Web Conference* (pp. 119–136). Springer, Cham. https://doi.org/10.1007/978-3-030-00668-6_8

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies, 1*(1), 428–444. https://doi.org/10.1162/qss_a_00023

Ritchie, A., Robertson, S., & Teufel, S. (2008). Comparing citation contexts for information retrieval. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 213–222). https://doi.org/10.1145/1458082.1458113

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

Schneider, J., Ye, D., Hill, A. M., & Whitehorn, A. S. (2020). Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. *Scientometrics, 125*(3), 2877–2913. https://doi.org/10.1007/s11192-020-03631-1

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

Zhang, Y., Cai, X., Fry, C. V., Wu, M., & Wagner, C. S. (2021). Topic evolution, disruption and resilience in early COVID-19 research. *Scientometrics, 126*(5), 4225–4253. https://doi.org/10.1007/s11192-021-03946-7