
USING ALTMETRICS FOR DETECTING IMPACTFUL RESEARCH IN QUASI-ZERO-DAY TIME-WINDOWS: THE CASE OF COVID-19

A PREPRINT

Erik Boetto

DIBINEM, University of Bologna, Italy
E-mail: erik.boetto@gmail.com
ORCID: 0000-0002-2509-159X

Maria Pia Fantini

DIBINEM, University of Bologna, Italy
E-mail: mariapia.fantini@unibo.it
ORCID: 0000-0002-3257-6552

Aldo Gangemi

STLab, ISTC-CNR, Rome, Italy
E-mail: aldo.gangemi@cnr.it
ORCID: 0000-0001-5568-2684

Davide Golinelli

DIBINEM, University of Bologna, Italy
E-mail: davide.golinelli@unibo.it
ORCID: 0000-0001-7331-9520

Manfredi Greco

DIBINEM, University of Bologna, Italy
E-mail: manfredi.greco@studio.unibo.it
ORCID: 0000-0001-6886-0160

Andrea Giovanni Nuzzolese✉*

STLab, ISTC-CNR, Rome, Italy
E-mail: andrea.nuzzolese@cnr.it
ORCID: 0000-0003-2928-9496

Valentina Presutti

LILEC, University of Bologna, Italy
E-mail: valentina.presutti@unibo.it
ORCID: 0000-0002-9380-5160

Flavia Rallo

DIBINEM, University of Bologna, Italy
E-mail: flavia.rallo@studio.unibo.it

April 15, 2020

ABSTRACT

Background. On December 31st 2019, the World Health Organization (WHO) China Country Office was informed of cases of pneumonia of unknown etiology detected in Wuhan City. The cause of the syndrome was a new type of coronavirus isolated on January 7th 2020 and named Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2). SARS-CoV-2 is the cause of the coronavirus disease 2019 (COVID-19). Since January 2020 an ever increasing number of scientific works have appeared in literature. Identifying relevant research outcomes at very early stages is challenging.

Objective. In this work we use COVID-19 as a use-case for investigating: (i) which tools and frameworks are mostly used for early scholarly communication; (ii) to what extent altmetrics can be used to identify potential impactful research in tight (i.e. *quasi-zero-day*) time-windows.

Approach. A literature review with rigorous eligibility criteria is performed for gathering a sample composed of scientific papers about SARS-CoV-2/COVID-19 appeared in literature in the tight time-window ranging from January 15th 2020 to February 24th 2020. This sample is used for building a knowledge graph that represents the knowledge about papers and indicators formally. This knowledge graph feeds a data analysis process which is applied for experimenting with altmetrics as impact indicators.

Results. We find moderate correlation among traditional citation count, citations on social media, and mentions on news and blogs. This suggests there is a common intended meaning of the citational acts associated with aforementioned indicators. Additionally, we define a method, i.e. the *Compre-*

*A.G. Nuzzolese led the work presented in this paper. The authors are sorted alphabetically as they equally contributed to this work.

hensive Impact Score (CIS), that harmonises different indicators for providing a multi-dimensional impact indicator. CIS shows promising results as a tool for selecting relevant papers even in a tight time-window.

Conclusions. Our results foster the development of automated frameworks aimed at helping the scientific community in identifying relevant work even in case of limited literature and observation time.

1 Introduction

A *zero-day* attack is a cyber attack exploiting a vulnerability (i.e. zero-day vulnerability) of a computer-software that is either unknown or it has not been disclosed publicly Bilge and Dumitras (2012). There is almost no defense against a zero-day attack. In fact, according to Bilge and Dumitras (2012), while the vulnerability remains unknown, the software affected cannot be patched and anti-virus products cannot detect the attack through signature-based scanning.

On December 31st 2019, the World Health Organization (WHO) China Country Office was informed of cases of pneumonia of unknown etiology detected in Wuhan City (Hubei Province, China), possibly associated with exposures in a seafood wholesale market in the same city². The cause of the syndrome was a new type of coronavirus isolated on January 7th 2020 and named Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2). Formerly known as the 2019 novel coronavirus (2019-nCoV), SARS-CoV-2 is a positive-sense single-stranded RNA virus that is contagious among humans and is the cause of the coronavirus disease 2019, hereinafter referred to as COVID-19 Gorbalenya (2020). Borrowing cyber security terminology, COVID-19 is a zero-day attack where the target system is the human immune system and the attacker is SARS-CoV-2. The human immune system has no specific defense against SARS-CoV-2. Being SARS-CoV-2 a new type of virus, there is no immunity provided by either natural or artificial immunity (i.e. antibodies or vaccines) humans can rely on. In the last three months, since the virus was first identified as a novel coronavirus in January 2020, an ever increasing number of scientific works have appeared in literature. Identifying relevant research outcomes at very early stages is utmost important for guiding the scientific community and governments in more effective research and decisions, respectively. However, traditional methods for measuring the relevance and impact of research outcomes (e.g. citation count, impact factor, etc.) might be ineffective due to the extremely narrow observation window currently available. Notoriously, indicators like citation count or impact factor require broader observation windows (i.e. few years) to be reliable Lehmann et al. (2008). Altmetrics might be valid tools for measuring the impact in quasi-zero-day time-window. Altmetrics³ have been introduced by Priem et al. (2012) as the study and use of scholarly impact measures based on activity in online tools and environments. The term has also been used to describe the metrics themselves. COVID-19 pandemic offers an extraordinary playground for understanding inherent correlation between impact and altmetrics. In fact, for the first time in human history, we are facing a pandemic, which is described, debated, and investigated in real time by the scientific community via conventional research venues (i.e. journal papers), jointly with social and on-line media.

In this work we investigate the following research questions:

- *RQ1*: Which are the platforms, systems and tools mostly used for early scholarly communication?
- *RQ2*: How is it possible to use altmetrics for automatically identifying candidate impactful research works in quasi-zero-day time-window?

For answering aforementioned research questions we carry out an experiment by using a sample of 212 papers on COVID-19. This sample has been collected by means of a rigorous literature review.

The rest of the paper is organised in the following way: Section 2 presents related work; Section 3 describes the material and method used for the experiments; Section 4 presents the data analysis we perform and the results we record; Section 5 discusses the results; finally, Section 6 presents our conclusions and future work.

2 Related work

An ever increasing amount of research work has investigated the role of altmetrics in measuring impact since they have been introduced by Priem et al. (2012).

²WHO - World Health Organization. Situation report - 1, Novel Coronavirus (2019-nCoV), January 21st 2020 https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4.

³The altmetrics manifesto: <http://altmetrics.org/manifesto/>

Correlation among indicators. Much research focuses on finding a correlation between altmetrics and traditional indicators. The rationale behind these works is based on the assumption that traditional indicators have been extensively used for scoring research works, and measuring their impact. Hence, their reliability is accepted by practice. Works such as Li and Thelwall (2012); Li et al. (2012); Bar-Ilan (2012); Thelwall et al. (2013); Sud and Thelwall (2014); Nuzzolese et al. (2019) follow this research line. These studies record moderate agreement (i.e. ~ 0.6 with Spearman correlation coefficient) with specific sources of altmetrics, i.e. Mendeley and Twitter. According to Thelwall (2018) and Nuzzolese et al. (2019), Mendeley is the on-line platform that provides an indicator (i.e. the number of Mendeley readers) that correlates well with citation counts after a time period consisting of few years. The meta-analysis conducted by Bornmann (2015) confirms this result, i.e. the correlation with traditional citations for micro-blogging is negligible, for blog counts it is small, and for bookmark counts from online reference managers, it is medium to large. Nevertheless, none of those studies take into account the key property of altmetrics, i.e. that they emerge quickly Peters et al. (2014). Hence, altmetrics should be used for measuring impact at very early stages, as soon as a topic emerges or a set of research works appear in literature. As a consequence, we use a tight time scale (i.e. quasi-zero-day time-window) for carrying out our analysis.

Altmetrics and research impact. The analysis of altmetrics with respect to research evaluation frameworks has been carried out by Wouters et al. (2015); Ravenscroft et al. (2017); Bornmann and Haunschild (2018); Nuzzolese et al. (2019). More in detail, Wouters et al. (2015) uses the Research Excellence Framework (REF) 2016, i.e. the reference system for assessing the quality of research in UK higher education institutions, for mining possible correlation among different metrics. The analysis is based on different metrics (either traditional or alternative) and research areas, and its outcomes converge towards limited or no correlation. Ravenscroft et al. (2017) finds very low or negative correlation coefficients between altmetrics provided by Altmetric.com and REF scores concerning societal impact published by British universities in use case studies. The aim of the analysis carried out by Bornmann and Haunschild (2018) is twofold. Bornmann and Haunschild (2018) investigates the correlation between citation counts and the relationship between the dimensions and quality of papers using regression analysis on post-publication peer-review system of F1000Prime assessments. Such a regression analysis shows that only Mendeley readers and citation counts are significantly related to quality. Finally, Nuzzolese et al. (2019) uses data from the Italian National Scientific Qualification (NSQ). The results show good correlation between Mendeley readers and citation count, and moderate accuracy for the automatic prediction of the candidates qualification at the NSQ by using independent settings of indicators as features for training a Naïve Bayes algorithm.

Some of the aforementioned works focuses on providing a comprehensive analysis investigating not only the correlation between traditional indicators and altmetrics, but also the correlation among the altmetrics themselves. However, all of them overlook the time constraint (i.e. a tight observation window), which is utmost important in our scenario.

3 Material and method

In this section we present the input data and the method used for processing such data. More in detail, we explain: (i) the approach adopted for carrying out the literature review focused on gathering relevant literature associated with the COVID-19 pandemic; (ii) the sources and the solution used for enriching the resulting articles with citation count as well as altmetrics; and (iii), finally, the method followed for processing collected data.

3.1 Literature review

The initial search was implemented on February 17th, 2020 in MEDLINE/Pubmed. The search query consists of the following search terms selected by the authors to describe the new pandemic: [coronavirus* OR Pneumonia of Unknown Etiology OR COVID-19 OR nCoV]. Although the name has been updated to SARS-CoV-2 by the International Committee on Taxonomy of Viruses⁴ on February 11th 2020, the search is performed by using the term “nCoV” because we presume that no one, between February 11th and 13th, would have used the term “SARS-COV-2”. Furthermore, the search is limited to the following time-span: from January 15th, 2019 to February 24th, 2020. Due to the extraordinary rapidity, with which scientific papers have been electronically published online (i.e. ePub), it may happen that some of these have indicated a date later than February 13th 2020 as publication date.

We rely on a two-stage screening process to assess the relevance of studies identified in the search. For the first level of screening, only the title and abstract are reviewed to preclude waste of resources in procuring articles that do not meet the minimum inclusion criteria. Titles and abstracts of studies initially identified are then checked by two independent investigators, and disagreement among reviewers are resolved through a mediator. Disagreement is resolved primarily

⁴<https://www.nature.com/articles/s41564-020-0695-z>

through discussion and consensus between the researchers. If consensus is not reached, another blind reviewer acts as third arbiter.

Then, we retrieve the full-text for those articles deemed relevant after title and abstract screening. A form developed by the authors is used to record meta-data such as publication date, objective of the study, publication type, study sector, subject matter, and data sources. Results, reported challenges, limitations, conclusions and other information are ignored as they are out of scope with respect to this study.

Eligibility criteria. Studies are eligible for inclusion if they broadly include data and information related to COVID-19 and/or SARS-CoV-2. Because of limited resources for translation, articles published in languages other than English are excluded. Papers that describe Coronaviruses that are not SARS-CoV-2 are excluded. There is no restriction regarding publication status. In summary, the inclusion criteria adopted are: (i) English language; (ii) SARS-CoV 2; (iii) COVID-19; (iv) Pneumonia of Unknown etiology occurred in China between December 2019 and January 2020. Instead, exclusion criteria are: (i) irrelevant titles not indicating the research topic; (ii) coronavirus not SARS-CoV 2; (iii) SARS, MERS, other coronavirus-related disease not COVID-19; (iv) not human diseases.

Data summary and synthesis. The data are compiled in a single spreadsheet and imported into Microsoft Excel 2010 for validation. Descriptive statistics were calculated to summarize the data. Frequencies and percentages are utilized to describe nominal data. In next section (cf. Section 3.2) we report statistics about collected papers.

3.2 Data processing workflow

Selected papers resulting from the literature review are used as input of the data processing workflow. The latter allows us to automatically gather quantitative bibliometric indicators and altmetrics about selected papers and to organise them in a structured format consisting of a knowledge graph. Figure 1 shows the number of papers in the sample grouped by publication date.

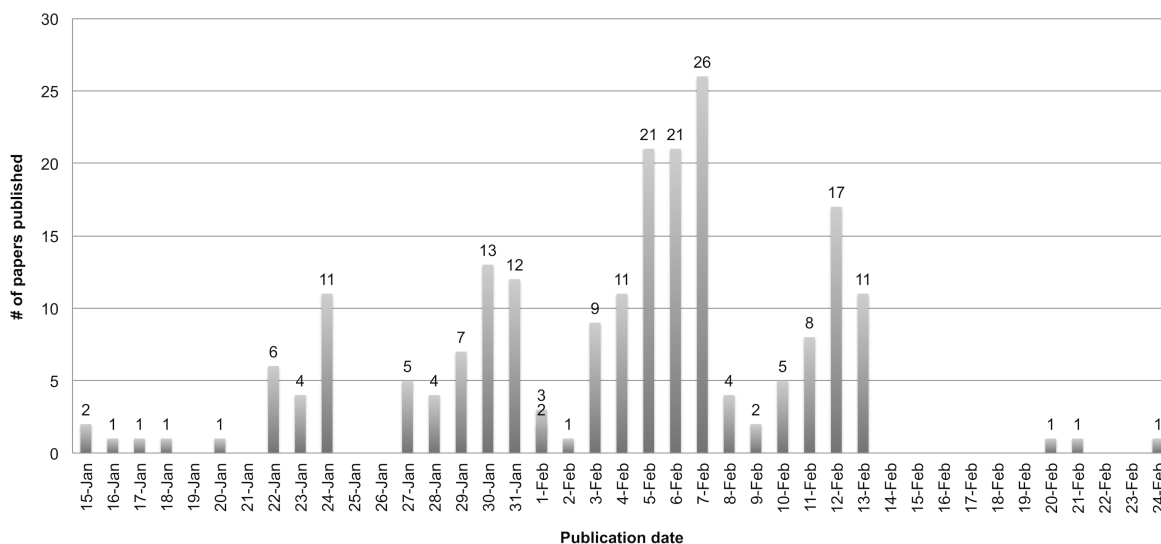


Figure 1: Number of papers per publication date in the time-window ranging from January 15th, 2020 to February 24th, 2020.

The workflow is based on an extension of the one we presented in Nuzzolese et al. (2019). Figure 2 shows the workflow as an UML activity diagram. In the diagram: (i) gray rectangles represent activities (e.g. the rectangles labelled “DOI identification”); (ii) gray boxes represent activities input pins; and (iii) white boxes represent activities’ output pins. The first activity is the identification of DOIs associated with selected papers. This is performed by processing the spreadsheet resulting from the literature review (cf. Section 3.1). Such a spreadsheet contains an article for each row. In turn, for each row, we take into account the following columns: (i) the internal identifier used for uniquely identifying the article within the CSV, (ii) the authors, (iii) the paper title, and (iv) the DOI whenever possible.

We rely on the Metadata API provided by Crossref⁵ for checking available DOIs and retrieving missing ones. This API is queried by using the first author and the title associated with each article as input parameters. Crossref returns

⁵<https://github.com/CrossRef/rest-api-doc>

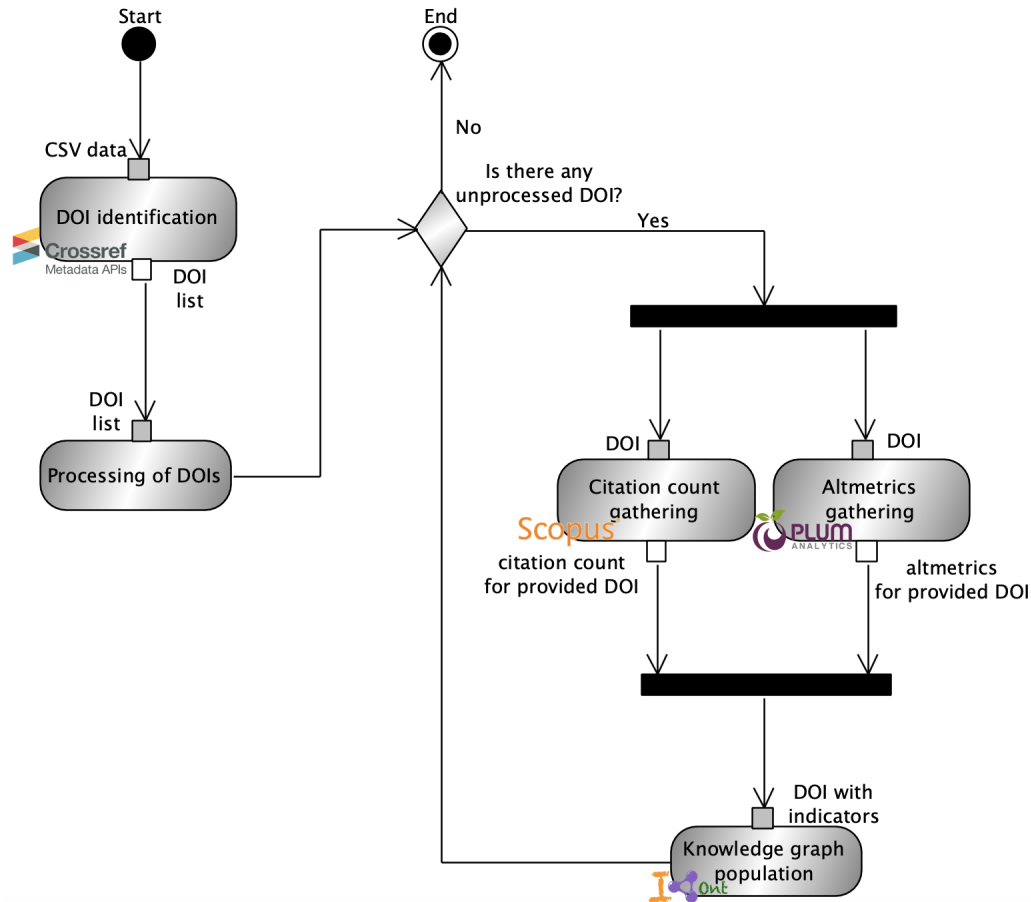


Figure 2: The UML activity diagram that graphically represents the workflow re-used and extended from (Nuzzolese et al. 2019) for the data processing activities.

the DOI that matches the query parameters as output. Whether a DOI is already available we first get the DOI from Crossref, then we check that the two DOIs (i.e. the one already available and one gathered from Crossref) are equal. In case the two DOIs are not equal we keep the DOI gathered from Crossref as valid. This criterion is followed in order to fix possible manual errors (e.g. typos) that would prevent the correct execution of subsequent actions of the workflow.

The output of the DOI identification activity is a list of DOIs which is passed as input to the second activity named “Processing of DOIs”. The latter iterates over the list of DOIs and selects them one by one. This operation allows other activities to gather information about citation count and altmetrics by using the DOI as the key for querying dedicated web services. The processing of DOIs proceeds until there is no remaining unprocessed DOI in the list (cf. the decision point labelled “Is there any unprocessed DOI?” in Figure 2).

The activities “Citation count gathering” and “Altmetrics gathering” are carried out in parallel. Both accept a single DOI as input parameter and return the citation count and the altmetrics associated with such a DOI, respectively. The citation count gathering relies on the API provided by Scopus⁶. We use Scopus as it is used by many organisations as the reference service for assessing the impact of research from a quantitative perspective (e.g. citation count, *h*-index, and impact factor). For example, the Italian National Scientific Habilitation⁷ (ASN) uses Scopus for defining threshold values about the number of citations and *h*-index scores that candidates to permanent positions of Full and Associate Professor in Italian universities should exceed. The altmetrics gathering activity is based on Plum Analytics⁸ (PlumX), which is accessed through its integration in the Scopus API⁹. We use PlumX among the variety of altmetrics

⁶https://dev.elsevier.com/tecdoc_cited_by_in_scopus.html

⁷<https://www.anvur.it/en/activities/asn/>

⁸<https://plumanalytics.com/learn/about-metrics/>

⁹<https://dev.elsevier.com/documentation/PlumXMetricsAPI.wadl>

providers (e.g. Altmetric.com or ImpactStory) as, according to Peters et al. (2014), it is the service that registers the most metrics for the most platforms. Additionally, in our previous study Nuzzolese et al. (2019), we found that PlumX is currently the service that covers the highest number of research work ($\sim 52.6M^{10}$) if compared to Altmetric.com ($\sim 5M^{11}$) and ImpactStory ($\sim 1M^{12}$). PlumX provides three different levels of analytics consisting of (i) the category, which provides a global view across different indicators that are similar in semantics (e.g. the number alternative citations of a research work on social media); (ii) the metric, which identifies the indicator (e.g. the number of tweets about a research work); (iii) and the source, that basically allows to track the provenance of an indicator (e.g. the number of tweets on Twitter about a research work). Hereinafter we refer to these levels as the category-metric-source hierarchy. Table 1 summarises the categories provided by PlumX by suggesting an explanation for each of them. A more detailed explanation about the categories, metric, and sources as provided by PlumX is available on-line¹³.

Table 1: The categories provided by PlumX along with an explanation about their semantics.

Category	Explanation	Category	Explanation
Usage	A signal that anyone is reading an article or otherwise using a research.	Captures	An indication that someone wants to come back to the work.
Mentions	Number of mentions retrieved in news articles or blog posts about research.	Social Media	The number of mentions included in tweets, Facebook likes, etc. that reference a research work.

Once the information about the citation count and altmetrics for an article is available, it is used for populating a knowledge graph in the activity labelled “Knowledge graph population”. The knowledge graph is represented as RDF and modelled by using the Indicators Ontology (I-Ont) Nuzzolese et al. (2018). I-Ont is an ontology for representing scholarly artefacts (e.g. journal papers) and their associated indicators, e.g. citation count or altmetrics such as the number of readers on Mendeley. I-Ont is designed as an OWL¹⁴ ontology and was originally meant for representing indicators associated with the papers available on ScholarlyData. ScholarlyData¹⁵ Nuzzolese et al. (2016) is the reference linked open dataset of the Semantic Web community about papers, people, organisations, and events related to its academic conferences. The resulting knowledge graph, hereinafter referred to as COVID-19-KG, is available on Zenodo¹⁶ for download. Table 2 reports the statistics recorded for the metric categories stored into the knowledge graph. We do not report statistics on minimum values as they are meaningless being them 0 for all categories.

Figure 3 shows the distribution of the indicators for each available metric. Namely, Figure 3a shows the distribution of the citation count over articles; Figure 3b shows the distribution of Shares, Likes & Comments, and Tweets (Social Media category); Figure 3c shows the distribution of Blog Mentions, News Mentions, Q&A Site Mentions, and References (Mentions category); and Figure 3d shows the distribution of Readers (Captures category) over articles. For each graphic line plotted in Figures 3a-d we report the DOIs associated with the papers that record the highest indicator value for the specific category. That is: (i) 10.1016/S0140-6736(20)30183-5 (appeared in The Lancet) for citation count, mentions, and social media; (ii) 10.1016/j.ijid.2020.01.009 (appeared in the International Journal of Infectious Diseases) for captures; and (iii) 10.1056/NEJMoa2001191 (appeared in the New England Journal of Medicine) for usage.

Finally, Figure 4 shows, for each category, the number of papers that count at least an indicator a given category. We provide for each category their underlying metrics and sources. The number of articles recorded at category level is not the sum of those reported at metric level. In fact, this number is the cardinality (i.e. size) of the intersection among

¹⁰Data retrieved from <https://plumanalytics.com/learn/about-metrics/coverage/> on March 2020.

¹¹Data retrieved from https://figshare.com/articles/Altmetric_the_story_so_far/2812843/1 on March 2020.

¹²Data retrieved from <https://twitter.com/Impactstory/status/731258457618157568> on March 2020.

¹³<https://plumanalytics.com/learn/about-metrics/>

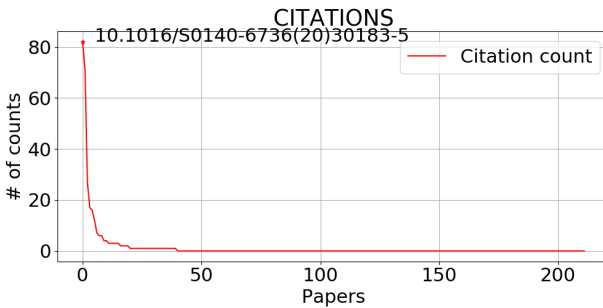
¹⁴<https://www.w3.org/TR/owl2-overview/>

¹⁵<https://w3id.org/scholarlydata>

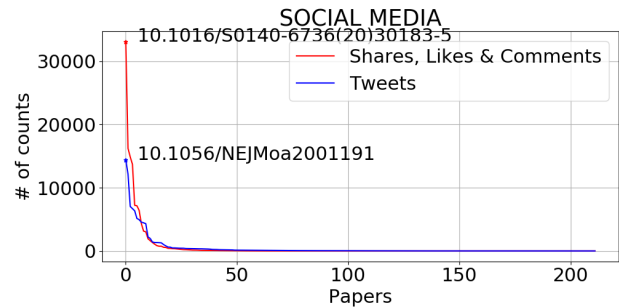
¹⁶<https://doi.org/10.5281/zenodo.3748694>

Table 2: The statistics recorded for metric categories including the citation count.

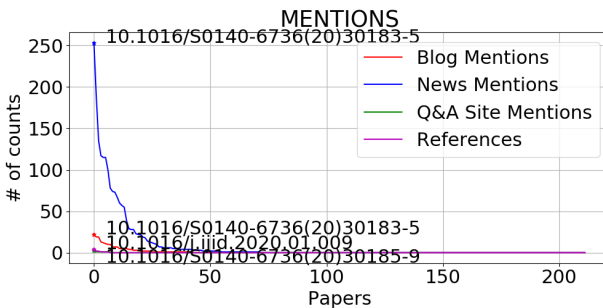
Category	Metric	Max	Mean	Median
Citations	Citation Count	82	1.63	0
	Total	82	1.63	0
Captures	Readers	161	1.69	0
	Total	161	1.69	0
Mentions	Blog Mentions	22	0.95	0
	News Mentions	253	8.95	0
	Q&A Site Mentions	3	0.05	0
	References	4	0.06	0
	Total	277	10.0	0
Usage	Abstract Views	15	0.07	0
	Total	15	0.07	0
Social Media	Shares, Likes & Comments	33,043	582.06	0
	Tweets	14,409	457.36	29.5
	Total	45,197	1,250.34	36.5



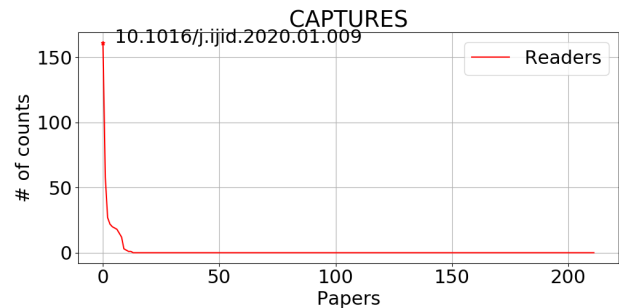
(a) Citations: Citation counts.



(b) Social Media: Shares, Likes & Comments, and Tweets.



(c) Mentions: Blog, News, Q&A Site, and References.



(d) Captures: Readers.

Figure 3: Distributions the number of indicators per paper.

the sets associated with underlying metrics. For instance, we record 157 as the indicator for the social media category, but we record 67 shares, likes & comments and 157 tweets. This means that 67 out of 157 papers are associated with both an indicator for shares, likes & comments and an indicator for tweets.

The workflow is implemented as a Python project and its source code is publicly available on GitHub¹⁷.

¹⁷<https://github.com/anuzzolese/covid-19-scientometrics>

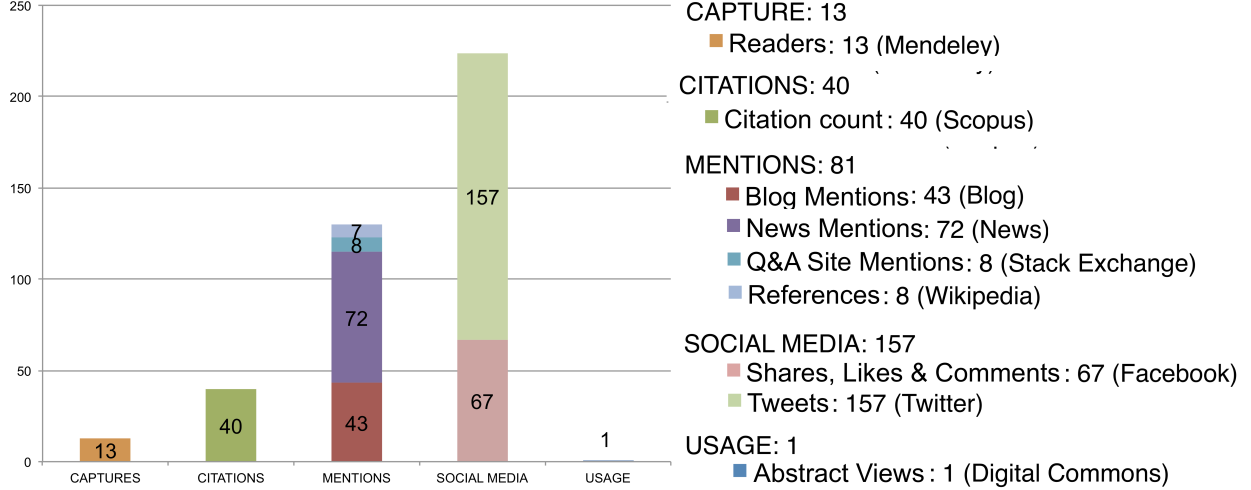


Figure 4: The number of articles that record at least one indicator value for each available category.

4 Data analysis

We design our experiment in order to address *RQ1* and *RQ2* by using COVID-19-KG. Hence, we first analyse the different indicators from a behavioural perspective, i.e. we want to investigate what are the indicators (social media, captures, etc.) and their underlying sources (e.g. Twitter, Mendeley, etc.) that perform better for scholarly communication in a narrow time-window (i.e. quasi-zero-day). Then, we analyse possible methods for identifying candidate impactful research work by relying on available indicators.

4.1 Behavioural perspective

In order to investigate the behaviour of collected indicators we set up an experiment composed of two conditions: (i) we compute the density estimation for each indicator in the category-metric-source hierarchy first on absolute values, then on standardised values; and (ii) we analyse the correlation among indicators.

Density estimation. The density provides a value at any point (i.e. the value associated with an indicator for a given paper) in the sample space (i.e. the whole collection of papers with indicator values in COVID-19-KG). This condition is useful to understand what are possible dense areas for each indicator. The density is computed with the Kernel Density Estimation (KDE) Scott (2015) by using Gaussian kernels. We use the method introduced by Silverman (1986) to compute the estimator bandwidth. We remark that the bandwidth is a non-parametric way to estimate the probability density function of a random variable. We opt for Silverman (1986) as it is one of the most used methods at the state of the art for automatic bandwidth estimation. The KDE is performed first by using absolute values (i.e. the values we record by gathering citation count and altmetrics) as sample set. Then, it is performed by using standardised values as sample set. The former is meant to get the probability distribution for each indicator separately. However, each indicator provides values recorded on very different ranges (cf. Table 2). Hence, KDE resulting from those different indicators are not directly comparable. Accordingly, we standardise indicator values and we then perform KDE over them. Again, KDE is performed for each indicator and for each level of the category-metric-source hierarchy. Standardised values are obtained by computing z -scores as the ratio between the sample mean and the standard deviation. Equation 1 formalises the formula we use for computing z -scores.

$$z(p, i) = \frac{p_i - \mu_i}{\sigma_i} \quad (1)$$

In Equation 1: (i) p_i is the value of the indicator i recorded for the paper p ; (ii) μ_i represents the arithmetic mean computed over the set of all values available for the indicator i for all papers; and (iii) σ_i represents the standard deviation computed over the set of all values available for the indicator i for all papers.

Figure 5 shows the diagrams of the KDEs we record for each category. For citation counts (cf. Figure 5a) the most dense area has d ranging from ~ 0.13 and ~ 0.001 and comprises articles that have from 0 to ~ 16 traditional citations. For social media (cf. Figure 5b) the most dense area has d ranging from ~ 0.00023 and ~ 0.00001 and comprises articles that have from 0 to $\sim 6,000$ alternative citations on social media. For mentions (cf. Figure 5c) the most dense

area has d ranging from ~ 0.029 and ~ 0.0008 and comprises articles that have from 0 to ~ 80 mentions. For captures (cf. Figure 5d) we record as the most dense area the one having density d ranging from ~ 0.08 and ~ 0.001 and comprising articles that count from 0 to ~ 20 number of captures. We do not compute the KDE for the usage category as there is one article only in COVID-19-KG with a value for such an indicator (cf. Figure 4).

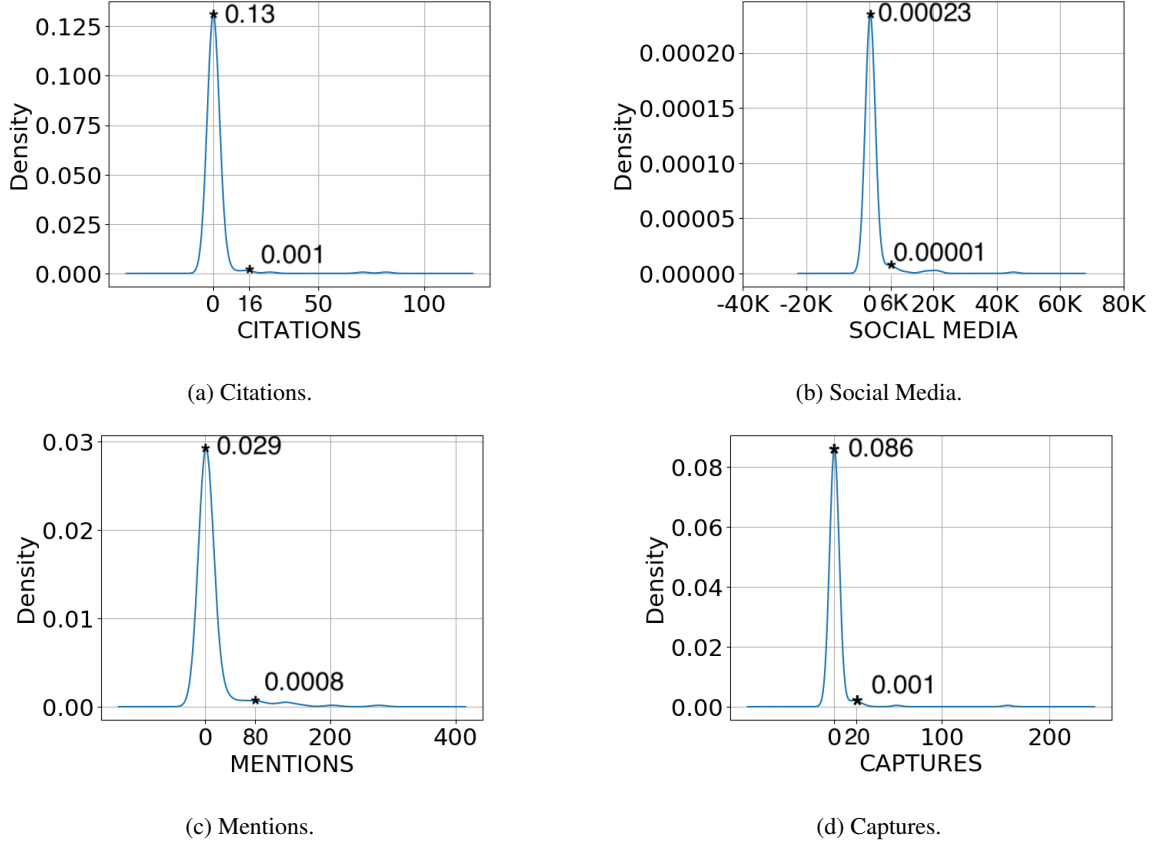


Figure 5: Diagrams of the kernel density estimations obtained for each category.

Instead, Figure 6 shows the KDE diagrams obtained with the standardised values. More in detail, Figure 6a and Figure 6b compare density estimation curves resulting from for the different categories and sources, respectively. We do not report KDE curves recorded for metrics as they are identical to those recorded for sources. This is due to the fact that there is a one-to-one correspondence between metrics and sources in COVID-19-KG, e.g. the Tweets metric has Twitter only among its sources. All most dense areas are those under the curve determined by d between ~ 1 and ~ 0.02 with values ranging from 0 to ~ 1 for selected indicators. This is recorded regardless of the specific level of the the category-metric-source hierarchy.

Correlation analysis. The correlation analysis aims at identifying similarities among different indicators both in their semantics and intended use on web platforms or social media (e.g. Twitter, Mendeley, etc.). This analysis repeats the experiment we carried out in Nuzzolese et al. (2019). We remind that in Nuzzolese et al. (2019) we used the papers extracted from the curricula of the candidates to the scientific habilitation process held in Italy for all possible disciplines as dataset. In the context of this work we narrow the experiment to a dataset with very peculiar boundaries in terms of (i) the topic (i.e. COVID-19) and (ii) the observation time-window (i.e. ranging from January 15th 2020 to February 24th 2020). As in Nuzzolese et al. (2019), we use the sample Pearson correlation coefficient (i.e. r) as the measure to assess the linear correlation between pairs of sets of indicators. The Pearson correlation coefficient is widely used in literature. It records correlation in terms of a value ranging from +1 to -1, where, +1 indicates total positive linear correlation, 0 indicates no linear correlation, and -1 indicates total negative linear correlation. For computing r , we construct a vector for each paper. The elements of a vector are the indicator values associated with its corresponding paper. We fill elements with 0 if an indicator is not available for a certain paper. The latter condition is mandatory in order to have vectors of equal size. In fact, r is computed by means of pairwise comparisons among vectors. The sample Pearson correlation coefficient is first computed among categories and then on sources by

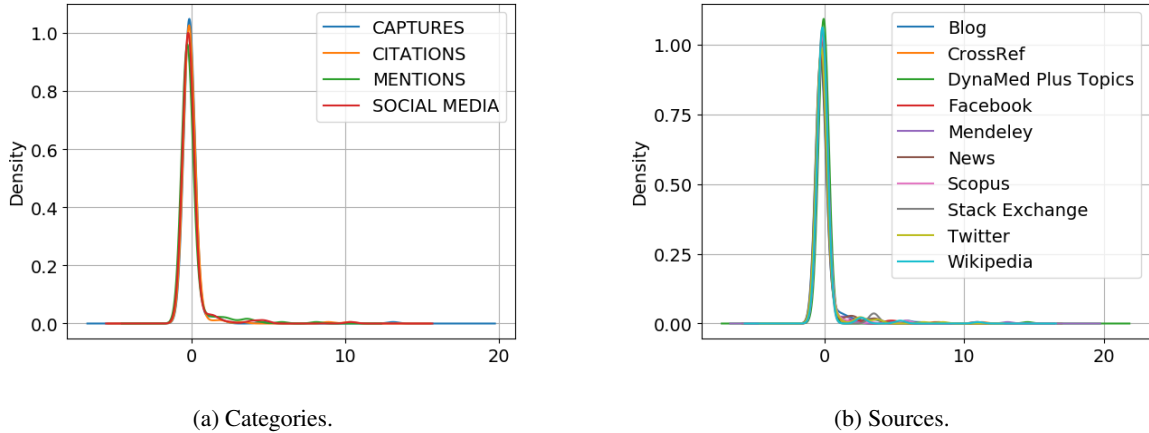


Figure 6: Diagrams of the KDEs computed on z -scores for categories (a) and sources (b).

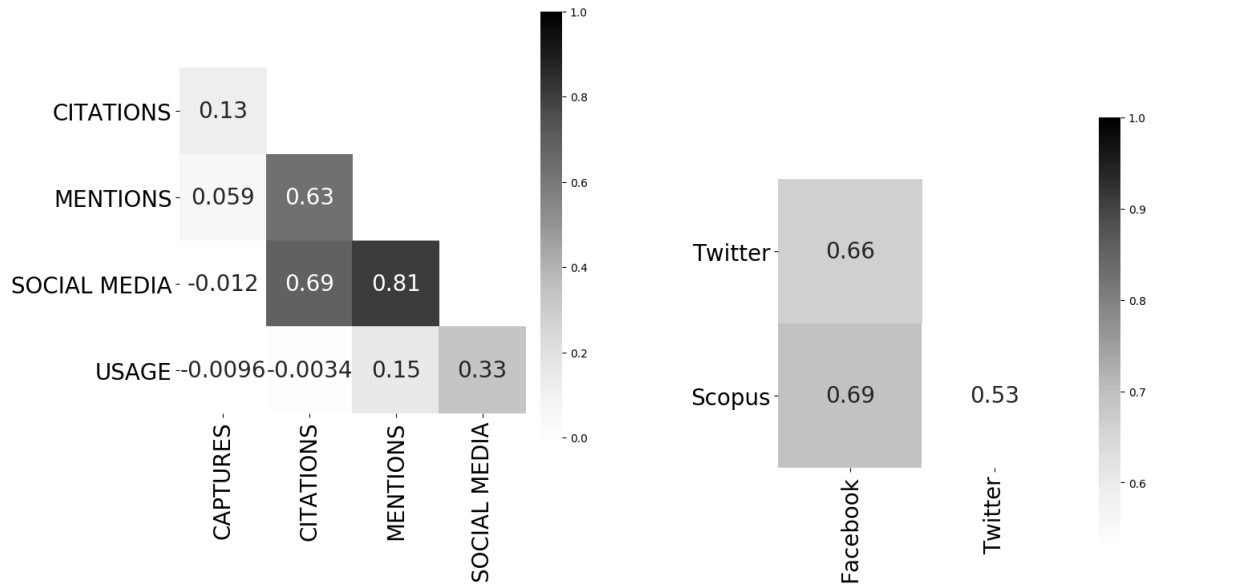
following the category-metric-source hierarchy as provided by PlumX. Again, we do not take into account the level of metrics as it is mirrored by the level of sources with a one-to-one correspondence. Additionally, r is investigated further only for those sources belonging to a category for which we record moderate correlation, i.e. $r > 0.6$. That is, we do not further investigate r if there is limited or no correlation at category level.

Figure 7 shows the confusion matrices resulting from the pairwise comparisons of the correlation coefficients. For categories (cf. Figure 7a) the highest correlation coefficients are recorded between: (i) mentions and citations, with $r=0.63$, statistical significance $p < 0.01$ (p -values are computed by using the Student’s t -distribution), and standard error $SE_r=0.04$; (ii) social media and citations, with $r=0.69$, $p < 0.01$, and $SE_r=0.04$; and (iii) social media and mentions, with $r=0.81$, $p < 0.01$, and $SE_r=0.03$. Figure 7b shows the confusion matrix for the sources associated with the social media and citations categories, i.e. Twitter and Facebook for social media and Scopus for citations. If we focus on cross-category sources only (i.e. we do not take into account moderate correlation coefficients recorded between sources associated with the same category) we record moderate correlation between Facebook and Scopus, with $r=0.69$, $p < 0.01$, and $SE_r=0.04$. Figure 7c shows the confusion matrix for the sources associated with the mentions and citations categories, i.e. News, Stack Exchange, and Wikipedia for mentions and Scopus for citations. The only cross-category sources associated with moderate correlation are News for mentions and Scopus for citations, with $r=0.63$, $p < 0.01$, and $SE_r=0.04$. Finally, Figure 7d shows the confusion matrix for the sources associated with the mentions and social media categories. In the latter we record $r > 0.6$ for the following cross-category sources: (i) Facebook and News, with $r=0.69$, $p < 0.01$, and $SE_r=0.04$; (ii) Facebook and Blog, with $r=0.62$, $p < 0.01$, and $SE_r=0.04$; (iii) Twitter and News, with $r=0.83$, $p < 0.01$, and $SE_r=0.03$; and (iv) Twitter and Blog, with $r=0.84$, $p < 0.01$, and $SE_r=0.03$.

4.2 Selecting candidate impactful papers

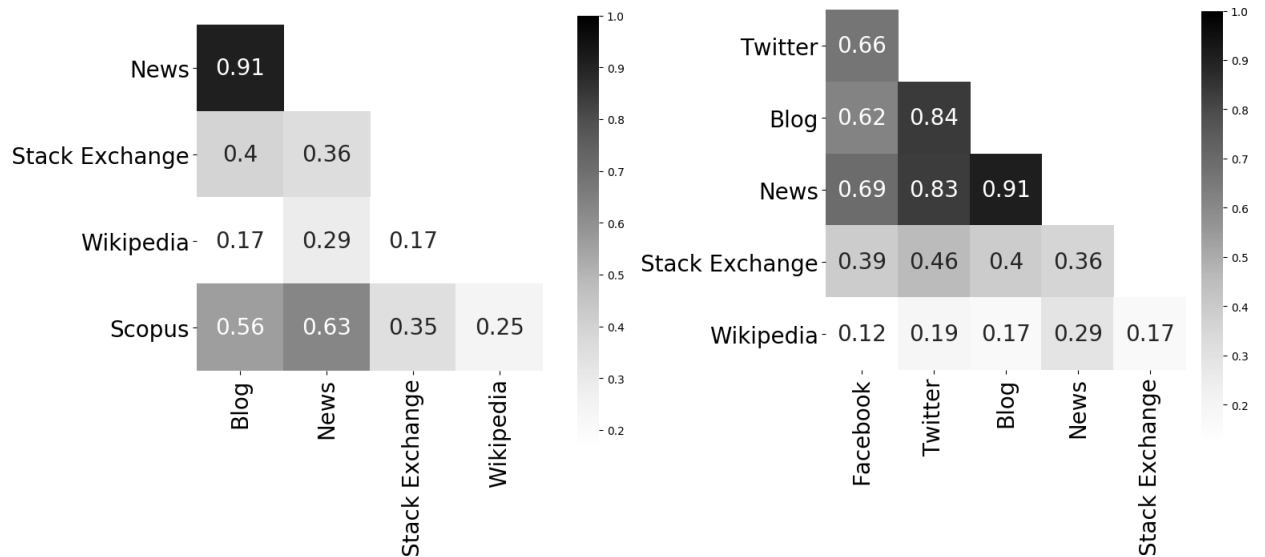
We then investigate how indicators can be used for selecting candidate impactful papers among those available in COVID-19-KG.

Geometric selection. First we rely on the result of the correlation analysis for selecting pairs of indicators that behave similarly. Hence, we use each pair for positioning papers on a Cartesian plane. Then we use such a positioning for defining a selection criterion. The axes of the Cartesian plane are the two indicators part of a pair. The axes values are the z -scores computed for each indicator (cf. Equation 1). We perform this analysis for the pairs (citations, social media), (citations, mentions), and (social media, mentions). We select these pairs only as they correlate better than others according to the correlation analysis (cf. Section 4.1). Furthermore, in COVID-19-KG citations, social media, and mentions are available for the most papers (cf. Figure 4). Figure 8 shows the results of this analysis. In order to draw a boundary around candidate impactful papers, we identify a threshold t for each category of indicators. We use the lower bound of the 95% quantiles, i.e. Q_{95} , as t . The quantiles are obtained by dividing the indicator values available for a given category (e.g. social media) COVID-19-KG into subsets of equal sizes. The lower bounds of the 95% quantiles recorded are 0.27, 1.11, and 1.75 for citations, social media, and mentions, respectively. For example, the Q_{95} for the citations category contains all that papers that count more than 0.27 citations each. We opt for 95% quantiles as they are selective. In fact, they allow us to gather the 5% papers in COVID-19-KG that record the highest value with respect to the selected indicator categories.



(a) Correlation among categories of indicators.

(b) Correlation among Twitter, Facebook, and Scopus citation count.



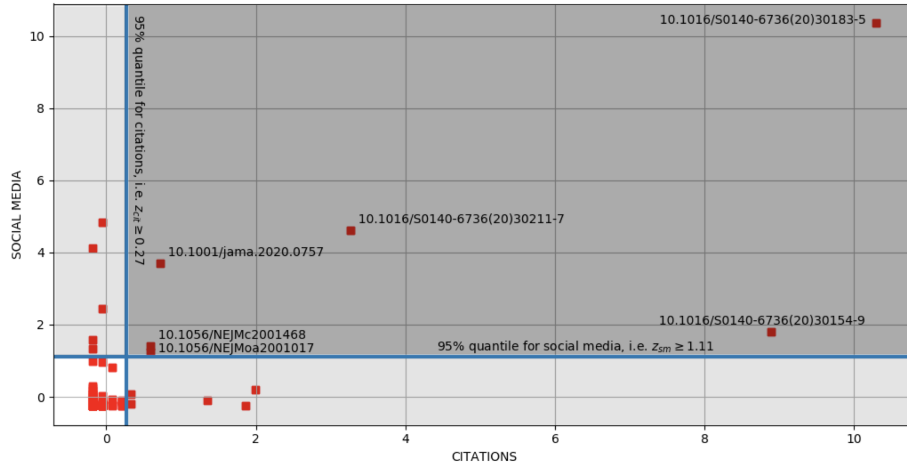
(c) Correlation among News, Stack Exchange, Wikipedia, and Scopus citation count.

(d) Correlation among News, Stack Exchange, Wikipedia, Twitter, and Facebook.

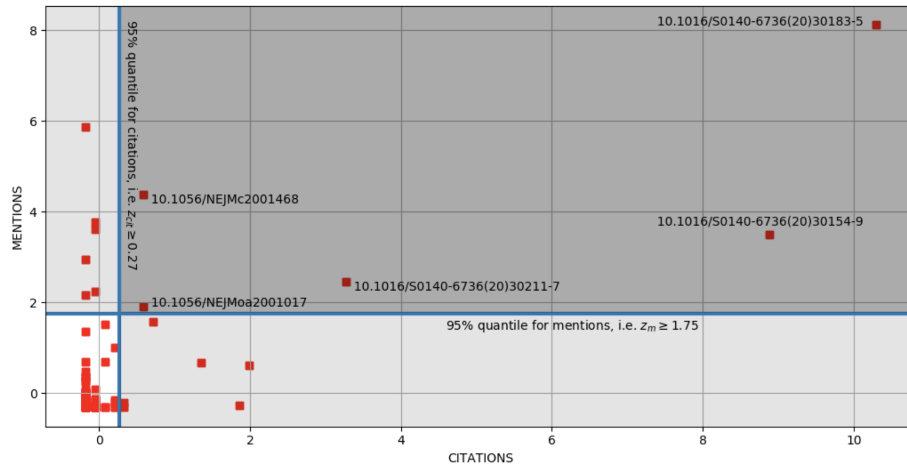
Figure 7: Correlation coefficients among indicators.

When we use citations and social media categories (we refer this combination to as $G_{c,s}$) as the axis of the Cartesian plane we record 6 papers whose indicator values are in Q_{95} of both categories (cf. Figure 8a). Instead, when we use citations and mentions categories (i.e. $G_{c,m}$) as axis we record 5 papers whose indicator values are in Q_{95} of both categories (cf. Figure 8b). Finally, when we use social media and mentions categories (i.e. $G_{s,m}$) as axis we record 9 papers whose indicator values are in Q_{95} of both categories (cf. Figure 7d).

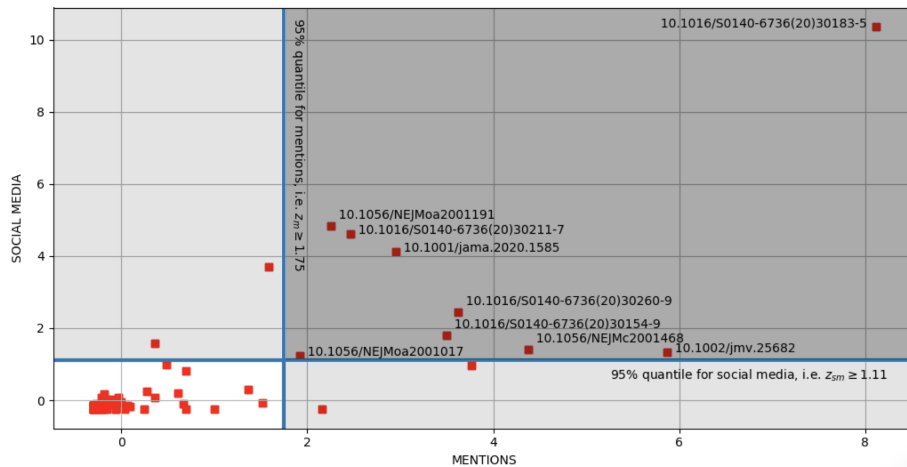
Comprehensive Impact Score. On top of the different indicators we compute, for each paper, a *Comprehensive Impact Score* (CIS). CIS aims at providing a multi-dimensional and homogeneous view over indicators which are different in quantities and semantics, i.e. CIS represents a unifying score over heterogeneous bibliometric indicators. A paper CIS is computed by first standardising the values associated with each indicator category (e.g. number of social media mentions, number traditional citations, etc.) and then averaging the resulting values. We use z -scores (cf.



(a) Citations and social media as axis.



(b) Citations and mentions as axis.



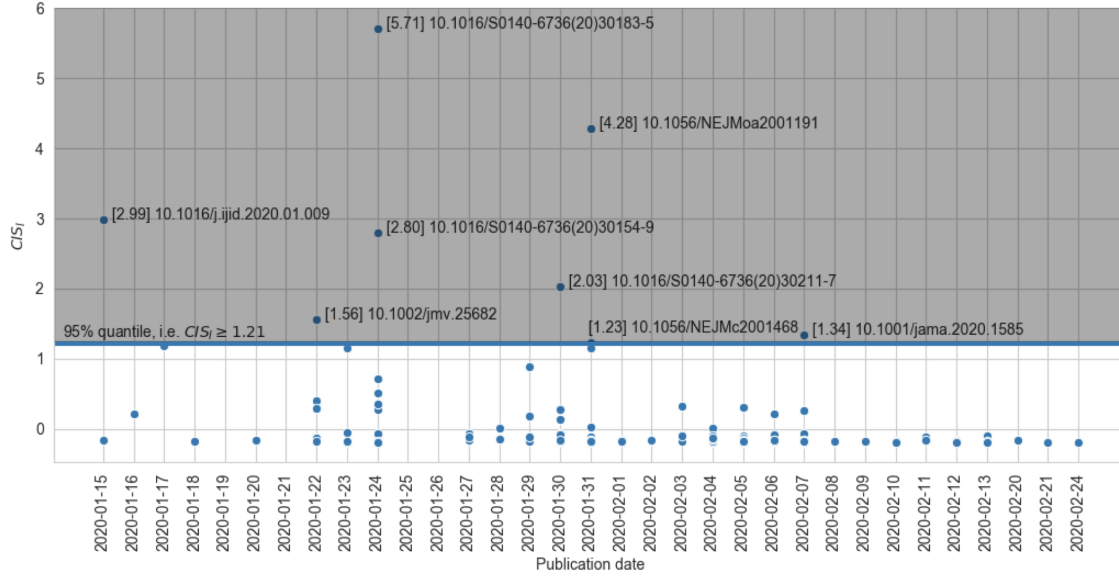
(c) Social media and mentions as axis.

Figure 8: Geometric spaces.

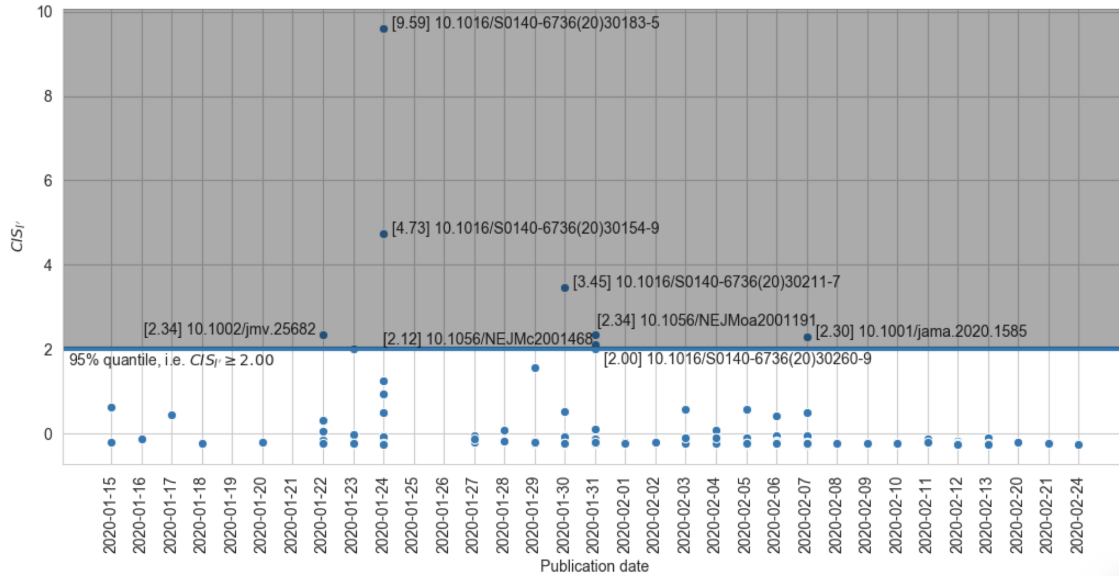
Equation 1) for obtaining standard values and the arithmetic mean for the average.

$$CIS(p) = \frac{\sum_{i \in I} z(p, i)}{\bar{I}} \quad (2)$$

In Equation 2: (i) p is a paper that belongs to the set of available papers in COVID-19-KG; (ii) i is an indicator that belongs to I , which, in turn, is the set of available indicators (e.g. citations, social media, etc.); and (iii) z is the function for computing z -scores as defined in Equation 1. Finally, we compute the 95% quantile on resulting CIS values. Again, the lower bound of the 95% quantile is used as threshold value (i.e. t) for identifying candidate impactful papers. We perform the selection of papers (i) first by using the whole set of available indicators (i.e. I) for computing CIS values, then (ii) by limiting the set of indicators to the categories of citations, social media, and mentions for computing CIS values. The limited set of indicators is referred to as I' and comprises the categories with highest correlation values.



(a) CIS computed on the all set of available indicators I .



(b) CIS computed on the set indicators I' limited to citations, social media, and indicators.

Figure 9: Selection of papers based on the Comprehensive Impact Score (CIS).

The lower bound of the 95% quantile for CIS values computed over I is 1.21. Such a lower bound for I' is 2. Figure 9 shows computed CIS values and selected papers. More specifically, Figure 9a shows CIS values computed on the whole set of indicators. Similarly, Figure 9b shows CIS values computed I' . Both figures present papers distributed according to their publication date. Furthermore, in those figures the threshold t is represented by the horizontal blue

line the cut off discarded papers (i.e. those points below the threshold line) from selected ones (i.e. those points above the threshold line). Table 3 summarises the candidate papers selected by using all different methods presented in this Section.

Table 3: Papers with their corresponding journal selected by using CIS_I , $CIS_{I'}$, $G_{c,s}$, $G_{c,m}$, and $G_{s,m}$.

Paper	Journal	CIS_I	$CIS_{I'}$	$G_{c,s}$	$G_{c,m}$	$G_{s,m}$
10.1016/j.ijid.2020.01.009	International Journal of Infectious Diseases		X			
10.1001/jama.2020.0757	JAMA			X		
10.1001/jama.2020.1585	JAMA	X	X			X
10.1002/jmv.25682	Journal of Medical Virology	X	X			X
10.1056/NEJMc2001468	New England Journal of Medicine	X	X	X	X	X
10.1056/NEJMoa2001017	New England Journal of Medicine			X	X	X
10.1056/NEJMoa2001191	New England Journal of Medicine	X	X			X
10.1016/S0140-6736(20)30260-9	The Lancet	X				X
10.1016/S0140-6736(20)30183-5	The Lancet	X	X	X	X	X
10.1016/S0140-6736(20)30154-9	The Lancet	X	X	X	X	X
10.1016/S0140-6736(20)30211-7	The Lancet	X	X	X	X	X

5 Discussion

The density estimation based on Gaussian kernels, i.e. KDE, shows that for COVID-19-KG all categories provide sparse indicators. However, if we analyse the density curves for individual metrics we observe clear patterns that characterise each indicator category uniquely. On one hand, we remark that, for the scope of this work, we investigate those patterns in order to understand how indicator categories behave by using the data coming from the COVID-19-KG. On the other hand, it is worth saying that patterns from KDE are specifically suitable for working in scenarios in which inference is required. For example, an algorithm might leverage learned patterns for implementing a classic binary classification task. This task might require to identify relevant papers from other samples with similar characteristics (e.g. similar time-window). A typical classification task might distinguish papers according to the relevant/not-relevant dichotomy. Accordingly, as future work, it would be interesting to associate KDE probability with impact categories, e.g. those emerging from the geometric space analysis or the CIS one. The KDE based on z -scores shows that density curves are mostly overlapped with each other, both at category and source level. Thus, we observe a similar citational behaviour once the indicator values are standardised as a shared pattern clearly emerges from their density curves. However, the KDE based on z -scores flattens differences that in terms of indicator values and their semantics, though implicit, are fairly evident. For example, social media counts range from 0 to 45,197 with 1,250.34 and 36.5 as mean and median, respectively, while citations ranges from 0 to 82 with 1.63 and 0 as mean and median, respectively (cf. Table 2). Numerical differences are captured by KDE if computed on individual categories (cf. Figure 5, but different semantic flavours among indicators are not).

We do not design any schema that formally captures the meaning of each indicator (e.g. an ontology like those proposed by D’Arcus and Giasson (2008) or Shotton (2010)). Nevertheless, we investigate if any correlation among any pair of indicators within the category-metric-source hierarchy can be interpreted as similarity in meaning. The correlation analysis suggests that citations, social media, and mentions identify a cluster of indicators that is used with a certain degree of consistency by citing entities. In Nuzzolese et al. (2019) we performed a similar analysis, and we recorded good correlation between citations and usage (i.e. the Mendeley readers). This different result should not misinterpreted. We are persuaded that the outcomes presented in Nuzzolese et al. (2019) are valid. Simply, they are obtained on a sample data which is inherently different from COVID-19-KG. Indeed, the data sample used in Nuzzolese et al. (2019) is much larger and wider in terms of papers contained (i.e. 833,116 against 212 in COVID-19-KG) and the time-window encompassing publication dates (i.e. many years against a month in COVID-19-KG), respectively. Accordingly, the research questions leading this work are different (cf. $RQ1$ and $RQ2$). As a matter of fact, the limited time-window is a mandatory requirement inherently related to the limited scientific literature we have since COVID-19 first appeared in human history (approximately December 2019). Based on our results we are happy to record social media (i.e. Twitter and Facebook) and mentions (limited to News and Blogs only) as valid tools for fast and early scholarly communication. This follows the visionary intuition of the altmetrics manifesto¹⁸ by Priem et al. (2012). The moderate occurrence of traditional citation counts with a certain positive (i.e. $r > 0.6$) and statistically significant (i.e. $p < 0.01$) correlation with altmetrics (i.e. social media and mentions) is, instead, a bit surprising. The

¹⁸<http://altmetrics.org/manifesto/>

statistics about usage (cf. Table 2), along with the results recorded for the KDE and the correlation analysis entitle us to claim that tweets on Twitter, shares and likes on Facebook, mentions on news and blogs, and citations in academic literature tracked by Scopus, are the channels that have been mostly used for early scholarly communication about COVID-19, i.e. this addresses *RQ1*.

The selection method based on the geometric space shows that mentions and citations are more selective than other pairs, when they are used together as axes of the Cartesian plane meant for positioning papers geometrically (cf. Table 3). In fact, by using them, we record a set of 5 candidate papers.

The intersection among the three sets of candidate papers gathered with $G_{c,s}$, $G_{c,m}$, and $G_{s,m}$ is equivalent to the set of 5 candidate papers obtained when using citations-mentions as axes (i.e. $G_{c,m}$). This suggests that selection of those 5 papers is reliable even if we have no evidence of its exhaustiveness. If we assess the selection based on the impact of the journal the papers have been published in, then we record good evidence about quality. In fact, both the New England Journal of Medicine and The Lancet are in the top-5 journal ranking on medicine according to SCImago¹⁹, with an SJR of 19.524 and 15.871, respectively. With regards to the exhaustiveness, the selection of candidate impactful papers is, in our opinion, an exploratory search task. According to White et al. (2005), exploratory search tasks are typically associated with undefined and uncertain goals. This means that identifying all possible impactful papers is nearly impossible. Hence, dealing with sub-optimal exhaustiveness is the practice in scenarios like these due to the inherent nature of the search problem itself.

The selection based on the Comprehensive Impact Score (CIS) overcomes the limitation of a two-dimensional space introduced when defining a selection method based on a Cartesian plane. Indeed, CIS is a multi-dimensional selection tool which is customisable in terms of the indicators used for performing the analysis. It is fairly evident (cf. Table 3) that both CIS_I and $CIS_{I'}$ share most of the papers identified by applying the two-dimensional geometric space selection with citations and mentions used as axes (referred to as $G_{c,m}$ in Table 3). CIS_I and $CIS_{I'}$ extend the set of selected papers returned by $G_{c,m}$ with 4 additional papers (cf. Table 3) published in The Lancet, the International Journal of Infectious Diseases ($SJR = 1.456$), JAMA ($SJR = 7.477$), the Journal of Medical Virology ($SJR = 0.966$). Among those additional papers only the one appeared in the Journal of Medical Virology is debatable both for the journal impact (i.e. $SJR = 0.966$) and its scientific relevance. In fact, in this paper the authors claim that two type of snakes, which are common in Southeastern China, are the intermediate hosts responsible for the “crossspecies” transmission of SARS-CoV-2 to humans. Subsequent genomic studies²⁰ Andersen et al. (2020) confirm crossspecies transmission, but they refute the theory of snakes being the intermediate hosts. However, the paper reporting the theory of snakes being the intermediate hosts has been largely (i) retweeted, shared, and liked on different social networks, and (ii) discussed and reported by many international newspapers worldwide. Thus it contributed to the massive “infodemic”²¹ about COVID-19, i.e. an over-abundance of information, either accurate or not, which makes it hard for people to find trustworthy sources and reliable guidance when they need it Zarocostas (2020). This infodemic is captured by altmetrics that by definition are fed by on-line tools and platforms. As a matter of fact, the paper about the theory of snakes being the intermediate hosts is selected among the candidates only by $G_{s,m}$, where both axes are altmetrics, and not by $G_{c,s}$ and $G_{c,m}$, where traditional citations are taken into account. This suggests a twofold speculation: (i) the scientometric community should handle altmetrics very carefully as they may lead to unreliable and debatable results; (ii) altmetrics are promising tools not only for measuring impact, but also to make unwanted scenarios (e.g. infodemic) emerge from the knowledge soup of scientific literature. We opt for the second. Nevertheless, further and more focused research is needed.

6 Conclusions and future work

In this work we investigate altmetrics and citation count as tools for detecting impactful research works in quasi-zero-day time-windows, as it is for the case of COVID-19. COVID-19 offers an extraordinary real-world case-study for understanding inherent correlation among impact and altmetrics. As mentioned in Section 1, for the first time in history, humankind is facing a pandemic, which is described, debated, and investigated in real time by the scientific community via conventional research venues (i.e. journal papers), and social and on-line media. The latter are the natural playground of altmetrics. Our case-study relies on a sample of 212 scientific papers on COVID-19 collected by means of a literature review. Such a literature review is based on a two-stage screening process used to assess the relevance of studies on COVID-19 appeared in literature from January 15th 2020 to February 24th 2020. This sample

¹⁹<https://www.scimagojr.com/journalrank.php?category=2701>

²⁰Andersen et al. (2020) is not part of the COVID-19 sample as it has been published on March 17th 2020, while the upper bound of time-window of the COVID-19 sample is February 24th 2020.

²¹<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf>

is used for constructing a knowledge graph, i.e. COVID-19-KG, modelled in compliance with the Indicators Ontology (i.e. I-Ont). COVID-19-KG is the input of our analysis aimed at investigating (i) behavioural characteristics of altmetrics and citation count and (ii) possible approaches for using altmetrics along with citation count for automatically identifying candidate impactful research works in COVID-19-KG. We find moderate correlation among traditional citation count, citations on social media, and mentions on news and blogs. This suggests there is a common intended meaning of the citational acts associated with these indicators. Additionally, we define the *Comprehensive Impact Score* (CIS) that harmonises different indicators for providing a multi-dimensional impact indicator. CIS shows to be a promising tool for selecting relevant papers even in a tight observation window. Possible future work include the use of CIS as a feature for predicting the results of evaluation procedures of academics as presented in works like Poggi et al. (2019). Similarly, further investigation is needed to mine the rhetorical nature of citational acts associated with altmetrics. The latter is a mandatory step for building tools such as Ciancarini et al. (2014) and Peroni et al. (2020). More ambitiously, future research focused on altmetrics and citations should go in the direction envisioned by Gil et al. (2014) and Kitano (2016), thus contributing to a new family of artificial intelligence aimed at achieving autonomous *discovery science*.

References

- K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of SARS-CoV-2. *Nature Medicine*, 1–3 (2020). doi:0.1038/s41591-020-0820-9
- J. Bar-Ilan, JASIST 2001–2010. *Bulletin of the Association for Information Science and Technology* **38**(6), 24–28 (2012). doi:10.1002/bult.2012.1720380607
- L. Bilge, T. Dumitras, Before we knew it: an empirical study of zero-day attacks in the real world., in *ACM Conference on Computer and Communications Security*, ed. by T. Yu, G. Danezis, V.D. Gligor (ACM, New York, 2012), pp. 833–844. ISBN 978-1-4503-1651-4. doi:10.1145/2382196.2382284
- L. Bornmann, Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics* **103**(3), 1123–1144 (2015). doi:10.1007/s11192-015-1565-y
- L. Bornmann, R. Haunschild, Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PloS one* **13**(5), 0197133 (2018). doi:10.1371/journal.pone.0197133
- P. Ciancarini, A.D. Iorio, A.G. Nuzzolese, S. Peroni, F. Vitali, Evaluating Citation Functions in CiTO: Cognitive Issues, in *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, ed. by V. Presutti, C. d’Amato, F. Gandon, M. d’Aquin, S. Staab, A. Tordai Lecture Notes in Computer Science, vol. 8465 (Springer, ???, 2014), pp. 580–594. doi:10.1007/978-3-319-07443-6_39
- B. D’Arcus, F. Giasson, *Bibliographic Ontology Specification*, Specification Document, 2008. <http://bibliontology.com/>
- Y. Gil, M. Greaves, J. Hendler, H. Hirsh, Amplify scientific discovery with artificial intelligence. *Science* **346**(6206), 171–172 (2014). doi:10.1126/science.1259439
- A.E. Gorbalenya, Severe acute respiratory syndrome-related coronavirus—the species and its viruses, a statement of the coronavirus study group. *BioRxiv* (2020). doi:10.1101/2020.02.07.937862
- H. Kitano, Artificial intelligence to win the nobel prize and beyond: Creating the engine for scientific discovery. *AI Magazine* **37**(1), 39–49 (2016). doi:10.1609/aimag.v37i1.2642
- S. Lehmann, A.D. Jackson, B. Lautrup, A quantitative analysis of indicators of scientific performance. *Scientometrics* **76**(2), 369–390 (2008). doi:10.1007/s11192-007-1868-8
- X. Li, M. Thelwall, F1000, Mendeley and Traditional Bibliometric Indicators, in *The 17th International Conference on Science and Technology Indicators*, ed. by Y.G. Archambault, V. Lariviere, 2012, pp. 541–551
- X. Li, M. Thelwall, D. Giustini, Validating online reference managers for scholarly impact measurement. *Scientometrics* **91**(2), 461–471 (2012). doi:10.1007/s11192-011-0580-x
- A.G. Nuzzolese, A.L. Gentile, V. Presutti, A. Gangemi, Conference Linked Data: The ScholarlyData Project, in *International Semantic Web Conference (2)*, ed. by P.T. Groth, E. Simperl, A.J.G. Gray, M. Sabou, M. Krtzsch, F. Lcu, F. Flock, Y. Gil Lecture Notes in Computer Science, vol. 9982, 2016, pp. 150–158. ISBN 978-3-319-46546-3. doi:10.1007/978-3-319-46547-0_16

- A.G. Nuzzolese, V. Presutti, A. Gangemi, P. Ciancarini, Extending ScholarlyData with research impact indicators, in *Workshop on Semantics, Analytics, Visualisation: Enhancing Scholarly Dissemination (SAVE-SD)*, Springer, 2018. doi:10.1007/978-3-030-01379-0_4
- A.G. Nuzzolese, P. Ciancarini, A. Gangemi, S. Peroni, F. Poggi, V. Presutti, Do altmetrics work for assessing research quality? *Scientometrics* **118**(2), 539–562 (2019). doi:10.1007/s11192-018-2988-z
- S. Peroni, P. Ciancarini, A. Gangemi, A.G. Nuzzolese, F. Poggi, V. Presutti, The practice of self-citations: a longitudinal study. *Scientometrics* **123**(1), 253–282 (2020). doi:10.1007/s11192-020-03397-6
- I. Peters, A. Jobmann, C.P. Hoffmann, S. Künne, J. Schmitz, G. Wollnik-Korn, Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie-praxis und forschung* **3** (2014)
- F. Poggi, P. Ciancarini, A. Gangemi, A.G. Nuzzolese, S. Peroni, V. Presutti, Predicting the results of evaluation procedures of academics. *PeerJ Computer Science* **5**, 199 (2019). doi:10.7717/peerj-cs.199
- J. Priem, P. Groth, D. Taraborelli, The altmetrics collection. *PloS one* **7**(11), 48753 (2012). doi:10.1371/journal.pone.0048753
- J. Ravenscroft, M. Liakata, A. Clare, D. Duma, Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PloS one* **12**(3), 0173152 (2017). doi:10.1371/journal.pone.0173152
- D.W. Scott, *Multivariate density estimation: theory, practice, and visualization* (John Wiley & Sons, ???, 2015)
- D. Shotton, CiTO, the Citation Typing Ontology. *Journal of Biomedical Semantics* **1**(Suppl 1), 6 (2010). doi:10.1186/2041-1480-1-s1-s6. <http://dx.doi.org/10.1186/2041-1480-1-s1-s6>
- B.W. Silverman, *Density estimation for statistics and data analysis*. Monographs on Statistics & Applied Probability, vol. 26 (Chapman and Hall, ???, 1986). ISBN 9780412246203
- P. Sud, M. Thelwall, Evaluating altmetrics. *Scientometrics* **98**(2), 1131–1143 (2014). doi:10.1007/s11192-013-1117-2
- M. Thelwall, Early Mendeley readers correlate with later citation counts. *Scientometrics* **115**(3), 1231–1240 (2018). doi:10.1007/s11192-018-2715-9
- M. Thelwall, S. Haustein, V. Larivière, C.R. Sugimoto, Do altmetrics work? Twitter and ten other social web services. *PloS one* **8**(5), 64841 (2013). doi:10.1371/journal.pone.0064841
- R.W. White, B. Kules, B.B. Bederson, Exploratory search interfaces: categorization, clustering and beyond: report on the XSI 2005 workshop at the Human-Computer Interaction Laboratory, University of Maryland. *SIGIR Forum* **39**(2), 52–56 (2005). doi:10.1145/1113343.1113356
- P. Wouters, M. Thelwall, K. Kousha, L. Waltman, S. de Rijcke, A. Rushforth, T. Franssen, The metric tide: Correlation analysis of REF2014 scores and metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management). London: Higher Education Funding Council for England (HEFCE) (2015). doi:10.13140/RG.2.1.3362.4162
- J. Zarocostas, How to fight an infodemic. *The Lancet* **395**(10225), 676 (2020). doi:10.1016/S0140-6736(20)30461-X