Metrics and Trends in Assessing the Scientific Impact

George Tsatsaronis[0000−0003−2116−2933]

Elsevier BV
Radarweg 29, 1043NX, Amsterdam, The Netherlands
g.tsatsaronis@elsevier.com

Abstract. The economy of science has been traditionally shaped around the design of metrics that attempt to capture several different facets of the impact of scientific works. Analytics and mining around (co-)citation and co-authorship graphs, taking into account also parameters such as time, scientific output per field, and active years, are often the fundamental pieces of information that are considered in most of the well adopted metrics. There are, however, many other aspects that can contribute further to the assessment of scientific impact, as well as to the evaluation of the performance of individuals, and organisations, e.g., university departments and research centers. Such facets may cover for example the measurement of research funding raised, the impact of scientific works in patented ideas, or even the extent to which a scientific work constituted the basis for the birth of a new discipline or a new scientific (sub)area. In this work we are going to present an overview of the most recent trends in novel metrics for assessing scientific impact and performance, as well as the technical challenges faced by integrating a plethora of heterogeneous data sources in order to be able to shape the necessary views for these metrics, and the novel information extraction techniques employed to facilitate the process.

Keywords: Scientific Impact · Metrics · Trends · Natural Language Processing · Machine Learning

1 Introduction

Measuring the impact of science has been traditionally approached by means of measuring the impact that scientific publications have. Though the notion of a scientific publication being the primary vessel of communicating science has its roots in the 17th century, the roots of scientometrics originate from the field of bibliometrics which appeared for the first time several centuries later; in fact many attribute the origin of the field to Paul Otlet, one of the founders of information science [17].

However, it is the way we interpret the word "*impact*", that has driven the research of scientometrics, almost ever since its birth. In this paper we will not attempt to add more interpretations to the existing ones; there is already a very comprehensive set of such interpretations, which have resulted into a number of academic and alternative metrics [19, 21]. The aim of this paper is to summarize the information needs of the different stakeholders served by scientometrics and to point to some recent research directions on how we can serve some of the unaddressed needs. Below we discuss the most important users served by the outcomes of scientometrics, as well as some of their most representative information needs. Eventually, serving all of these information needs entails combining state-of-the-art data analytics, data visualization, natural language processing, machine learning and information retrieval [12–14].

Researchers: The primary users of such metrics, with the major need being the awareness of their standing in their scientific fields. They also want to know the most important journals in their field of research, the most prominent researchers for collaborations, as well as the top universities and scientific (sub-)fields in their areas. Furthermore, they would like to know the trends, as well as the top funded areas, and the respective funders in their field, in order to look for funding opportunities.

Universities: Their overall, and field-specific, standing in the academic landscape is their primary need, which is used in turn for the national assessments and ranking. For their undergraduate and graduate study programs they need to be constantly aware of how the different scientific fields and trends evolve over time. Also, being aware of who the most prominent scientists are for each research field is important for shaping hiring plans. Monitoring of the funding landscape, funding trends and opportunities is also important as it affects the shaping of their research strategy.

Funders: Their most important information need is their ability to trace back the research outcomes of their grants, as well as the overall impact these brought to society. They are also interested in the top funded areas, the emerging scientific fields and trends, as well as in knowing the overall standing of universities and individual researchers, per field, which can be used, among other criteria, to the assessment of research grant proposals.

Journal Editors: They would like to know what are the most prominent researchers in the scientific scopes of their journals, as well as how that scope evolves over time. This helps editors manage the editorial board, and having potentially the top experts in the respective fields, included. Analysis of the trends might also lead to the creation of special issues, in order to give emphasis in the most recent impactful works. They also need to be aware of the overall standing of the journal in the journal's fields of research.

Reviewers: They need to be aware of the most important and impactful articles in their scientific field. An analysis of the standing/ranking of the different journals per field also helps assess and compare potentially relevant references

or material with impact, that is at the core of the research described in the reviewed article.
Publishers: The ability to monitor the trends across all fields of science, as well as an overview of the journals' rankings, top researchers, and universities are the primary information needs that publishers have from scientometrics.
Science Journalists: Bridging the gap between the research community and the rest of the society, science journalists have as primary information needs the impact of individual scientific articles. Trends, as well as journals' and universities' rankings are also very important.
Tax Payer: Tax payers often need to understand the scientific and societal impact of the research that was funded by state/public resources.
Global Community/General Public: For instance patients interested in understanding novel research on diseases, or, understanding context and authority (top institutions, journals, experts) and being able to distinguish the high quality research work among all the noisy information out there.

It is evident that many of the aforementioned information needs require the linking of multiple sources. For instance, being able to provide an analysis of the top funded research (sub)fields, entails the ability to annotate scientific articles with domain labels in different granularities, the capacity to automatically extract funding information from articles, as well as from the reports of the funders' research outcomes, and combine these pieces of information together. Further to that, if besides volume of funded articles the information need pertains to actual amounts in different currencies, then, in addition to the aforementioned sources, one would need to be able to scrape grants' information from funders' sites, and link the grants' metadata with the rest, to draw sums per field.

As complicated as it appears to be, the communities of natural language processing, machine learning, analytics and visualization combined already have the answers to the advanced techniques required to answer such complex information needs. In the remaining of the paper we will first provide an overview of the current best practices in measuring scientific impact (Section 2), as well as examples of novel, experimental technologies developed by *Elsevier*[1], in collaboration with research institutions and universities across the globe, to address the complex landscape of scientometrics, serving all of the aforementioned stakeholders (Section 3).

2 Approaches

The scientific impact in academia is primarily measured using citation-based metrics. The principle behind all of these metrics is to model how knowledge disseminates among scientists and their communities. There are also metrics that capture the impact of scientific works by looking outside academia, e.g., alternative metrics that examine social media, news articles and the attention that scientific works draw by the non-scientific public. In the following we give

[1] https://www.elsevier.com/

a high level overview of the most common such metrics used, and we conclude this section with some interesting experimental research works which utilize alternative views of this data. For a more thorough overview of existing metrics, the reader might wish to consult survey articles in the fields, e.g., [16].

2.1 Author-level Metrics

Some of the most common author-level metrics include the number of citations, the author's *h-index*, the i-10 index, and an incredibly large number of variations with increasing complexity (e.g., a comprehensive survey can be found at [22]), most often weighed with regards to the scientific field or portfolio of the author.

2.2 Article-level Metrics

Article-level metrics (ALMs) quantify the reach and impact of published research articles. Well established citation databases, such as *Scopus*[2][2], integrate data from various sources. For example, *Scopus* integrates the *PlumX Metrics*[3], which is a wide family of article-level metrics, along with traditional measures (such as citations), to present a richer and more comprehensive picture of an individual article's impact. Examples include citations, not only from other scientific articles, but also from clinical studies, patents and policies, usage (e.g., article downloads, views, video plays), captures (e.g., bookmarks, code forks), mentions (e.g., wiki mentions, news mentions), and social media (e.g., tweets).

2.3 Journal-level Metrics

At the journal level, one can compute some of the traditional metrics, e.g., h-index for the whole journal, or any of its variations. However, some additional metrics, with time bounds, have been more adopted for the assessment of a journal. CiteScore metrics for example, are a suite of indicators calculated from data in *Scopus*. At its basis, CiteScore averages the sum of the citations received in a given year to publications published in the previous three years, to the sum of publications in the same previous three years. The rest of the *CiteScore* metrics are calculated based on this indicator. The *SCImago Journal Rank* (*SJR*) is based on the concept of a transfer of prestige between journals via their citation links. Drawing on a similar approach to *Google*'s *PageRank*, *SJR* weights each incoming citation to a journal by the *SJR* of the citing journal, with a citation from a high-*SJR* source counting for more than a citation from a low-*SJR* source. Like *CiteScore*, *SJR* accounts for journal size by averaging across recent publications and is calculated annually. *Source Normalized Impact per Paper* (*SNIP*) is a sophisticated metric that intrinsically accounts for field-specific differences in citation practices. It does so by comparing each journal's citations per publication with the citation potential of its field, defined as the set of publications

[2] `http://scopus.com/`

[3] `https://plumanalytics.com/learn/about-metrics/`

citing that journal. *SNIP* therefore measures contextual citation impact and enables direct comparison of journals in different subject fields, since the value of a single citation is greater for journals in fields where citations are less likely, and vice versa. Last but not least, *Journal Impact Factor* (*JIF*) is calculated by *Clarivate Analytics* as the average of the sum of the citations received in a given year to a journal's previous two years of publications divided by the sum of citable publications in the previous two years.

2.4 Experimental Methods

The potential of working with the citation, co-citation, and co-authorship graphs in the field of scientometrics has given birth to a number of novel ideas, primarily by repurposing successful graph mining techniques. In many of such research works, e.g., [3, 18] the authors attempt to predict trends in the respective graphs, e.g., citations, collaborations, and in general how these graphs are going to evolve over time. Such methods enable detecting earlier impactful articles, as well as authors whose collaboration network and citations are growing fast (also known in the literature as *rising stars*). Lately, there is also attention in attempting to model the performance of universities and research institutions, and make predictions for their future state regarding funding, ranking and other factors, e.g. [20].

3 Filling the Information Needs Gap

In this section we are presenting three novel research directions that enable more granularity to some of the aforementioned metrics, and they also support addressing some of the information needs mentioned earlier, which the current metrics cannot serve.

3.1 Funding

Within the economy of the research market, funding bodies need to ensure that they are awarding funds to the right research teams and topics so that they can maximize the impact of the associated available funds. At the same time, funding organisations require public access to funded research adopting, for instance, the *US* Government's policy that *all federal funding agencies must ensure public access to all articles and data which result from federally-funded research*. As a result, institutions and researchers are required to report on funded research outcomes, and acknowledge the funding source and grants. In parallel, funding bodies should be in a position to trace back these acknowledgements and justify the impact and results of their research allocated funds to their stakeholders and the tax-payers alike. Researchers should also be able to have access to such information, which can help them make better educated decisions during their careers, and help them discover appropriate funding opportunities for their scientific interests, experience and profile.

This situation creates unique opportunities for publishers, and more widely, the affiliated industry, to coordinate and develop novel solutions that can serve funding agencies and researchers. A fundamental problem that needs to be addressed is, however, the ability to extract automatically the funding information from scientific articles, which can in turn become searchable in bibliographic databases. We have addressed this problem by developing a novel technology to automatically extract funding information from scientific articles [9], using natural language processing and machine learning techniques. The pipeline is carefully engineered to accept a scientific article as input in raw text format, and provide the detected funding bodies and associated grants as output annotations. For the engineering of the final solution we have exhaustively tested a number of state-of-the-art approaches for named entity recognition and information extraction. The advantage of the developed technology lies in its ability to learn how to combine a number of base classifiers, among which many are open source and publicly available, in order to create an *ensemble mechanism* that selects the best annotations from each approach.

The problem can be formulated as follows: given a scientific article as raw text input, denoted as T, the automated extraction of funding information from text translates in two separate tasks. First, identify all text segments $t \in T$, which contain funding information, and, second, process all the funding text segments t, in order to detect the set of the funding bodies, denoted as FB, and the set of grants, denoted as GR that appear in the text. Provided that there is training data available, the former problem can be seen as a binary text classification problem, where, given T and the set of all non-overlapping text segments t_i, such that the $\cup_i t_i = T$ (where $t_i \in T$), a trained binary classifier can decide for the class label of t_i, i.e., $C_{t_i} = 1$, if t_i contains funding information, or $C_{t_i} = 0$ if not. The latter task can be mapped to a named entity recognition (NER) problem, where given all t_i for which $C_{t_i} = 1$, the objective is to recognize within them all strings s, such that either $s \in FB$, i.e., it represents a funding body, or $s \in GR$, i.e., it represents a grant. There is a number of additional dimensions that one may consider in the formulation of this problem, such as additional entities like *Programs* or *Projects*, or detecting and labelling the funding relation between the funding bodies and the authors, e.g., *Monetary*, or *In-kind*. We argue that such a technology can be used in combination with existing metrics, to sufficiently address a significant portion of the funders' and researchers' information needs around funded articles and funding, respectively.

3.2 Colouring of Citations

As discussed earlier in the overview of the current most common scientometrics, impact is primarily quantified, and not necessarily qualified, e.g., by counting for example number of citations. These metrics have raised some criticism as they don't account for different qualitative aspects of the citations. Negative or self-citations [8] should be weighted in a different way compared, for example, to affirmative or methodological citations. The question of qualitative bibliometrics is, therefore, gaining more interest in literature and researchers are suggesting different approaches to the problem, e.g., [1].

The qualitative analysis of citations functions is not only important for bibliometrics purposes; it can also help researchers in their daily work. Browsing references and lists of cited works is a time consuming activity which can be made easier by automatically highlighting those aspects a scholar is looking for. This might be the case of a PhD student who is interested only in those works cited because they use the same methods of the experiment she is studying, or in those works cited because they agree on a specific theory. Having those specific papers highlighted with a simple click would save precious time from the daily routine of researchers. One of the first step in this direction is the delineation of a citation functions schema which works as a basis for an automatic citation characterisation tool. This is not an easy task considering the different features and aspects that one has to take into account. Despite the indisputable value of author's motivations for citation, these might not be the only characterizations a user is looking for, while surveying references and lists of citations. For this purpose, in collaboration with *University of Bologna*[4] we have conducted a study to assess which of these functions are deemed important by scholars [7], and we have further developed a deep machine learning approach that can automatically classify the type of each citation made in an article. The approach is based on the fusion of sentence embeddings, section type semantic encodings, main verb embeddings, and *SciCite*'s predictions [4], into a transformer-based model. As a result, citations can be actually qualified with this approach, and respective retrieval filters can be applied in production facing platforms, to filter on papers cited for specific reasons/intents.

3.3 Novelty and Trends

Elsevier's *Scival's Topics of Prominence*[5], provide a very comprehensive view of how science can be organized into topics, by creating a topic modeling which is primarily based on citations (e.g., [11]). Motivated by the interest that such mining and analysis attracts, we are also exploring novel ways of addressing the very important need of measuring trends and capturing new terminology appearing in the various scientific fields.

For this purpose, we have developed a deep learning approach [6], and a topic analysis-based approach [10], as research prototypes. Combined they can provide a thorough scanning of the latest, novel and influential terminology across all, or selected, scientific fields. The former approach learns feature representations from a target document (whose terminological novelty is to be inferred) with respect to the source document(s) using a *Convolutional Neural Network* (*CNN*), and is based on a recent sentence embedding paradigm [5]. We leverage their idea and create a representation of the relevant target document relative to the

[4] `https://www.unibo.it/en`

[5] `https://www.elsevier.com/solutions/scival/releases/topic-prominence-in-science`

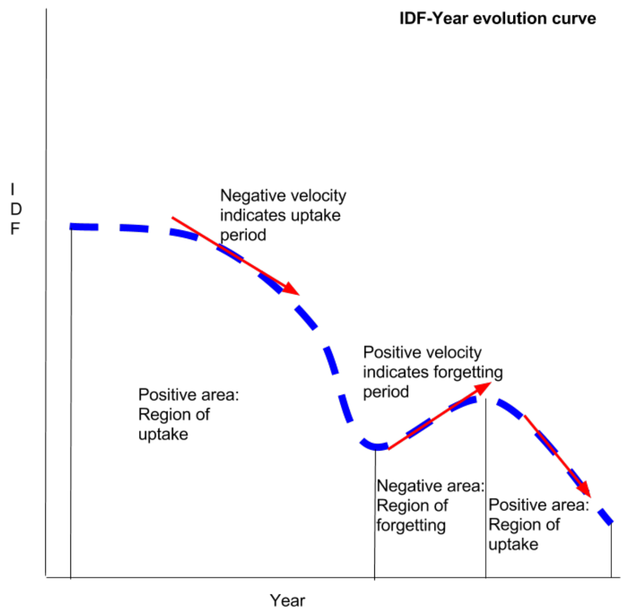

Fig. 1. Patterns of the *Inverse Document Frequency* (*IDF*) of a scientific topic over time, motivating the introduced measures of *topic attentionality* [10].

designated source document(s) and call it the *Relative Document Vector* (*RDV*). We can then train a *CNN* with the *RDV* of the target documents, and, finally, classify a document as terminologically novel or non-novel with respect to its source documents.

Next, we can apply the topic attentionality approach [10] in these documents, to extract specific novel terminology per area. The motivation behind this approach is to understand the velocity of the changes in the *Inverse Document Frequency* (*IDF*) of terms, as shown in Figure 1. At some point in time, the topic appears for the first time in the literature. Since it has not been discussed before, at that point in time its *IDF* score will be high. After that point in time, there might be a period where the topic acquires attention. During this period, its *IDF* score will be dropping, as the topic will be discussed more over time. During this period also, one can observe a negative velocity in the *IDF* curve, since the score is becoming gradually smaller. The area below such a negative velocity curve is in fact a positive area for the topic, as it describes the volume of the attention the topic is receiving; an attention which is gradually increasing. Further in time, the topic might be saturated by the research community, and then in the topic's *IDF* curve the reverse phenomenon might be observed: positive velocity *IDF* curve, since the topic is being discussed less over time from that point and on, meaning that it does not receive so much attention anymore.

The area under this positive velocity *IDF* curve is in fact a negative area of the topic, as it quantifies the volume of the attention the topic lost over time. In principle, these two patterns, namely negative *IDF* velocity (topic attracts attention) and positive *IDF* velocity (topic loses attention) might alternate for the same topic over time, and are the two main motifs of the *IDF* values of the topic measures over time. The ability to compute such metrics across all candidate novel terms, and across fields, can address sufficiently the problem of detecting (sub)field trends, and one could also trace back the origin/main contributors of the shaping of new areas. One can also notice the relation of this idea to the notion of delayed recognition in science as well [15].

4 Summary

In this paper we have provided an overview of the major users and recipients of scientometrics output and analyses, along with their most representative information needs. We have noted that there are still significant gaps in addressing these needs, and we have discussed a few directions that can add more clarity and granularity to existing metrics. The three directions, namely mining and linking funding information, qualifying citations and classifying citation intent, and detecting novelty and trends in scientific terminology, can enable the development of novel scientometrics, and can help close the gap by addressing the remaining information needs.

References

1. Abu-Jbara, A., Ezra, J., Radev, D.R.: Purpose and polarity of citation: Towards nlp-based bibliometrics. In: Vanderwende, L., III, H.D., Kirchhoff, K. (eds.) Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA. pp. 596–606. The Association for Computational Linguistics (2013)
2. Baas, J., Schotten, M., Plume, A.M., Côté, G., Karimi, R.: Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quant. Sci. Stud. **1**(1), 377–386 (2020)
3. Bai, X., Zhang, F., Lee, I.: Predicting the citations of scholarly paper. J. Informetrics **13**(1), 407–418 (2019)
4. Cohan, A., Ammar, W., van Zuylen, M., Cady, F.: Structural scaffolds for citation intent classification in scientific publications. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 3586–3596. Association for Computational Linguistics (2019)
5. Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A.: Supervised learning of universal sentence representations from natural language inference data. In: Palmer, M., Hwa, R., Riedel, S. (eds.) Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen,

Denmark, September 9-11, 2017. pp. 670–680. Association for Computational Linguistics (2017)
6. Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Tsatsaronis, G., Chivukula, S.S.S.K.: Novelty goes deep. A deep neural solution to document level novelty detection. In: Bender, E.M., Derczynski, L., Isabelle, P. (eds.) Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018. pp. 2802–2813. Association for Computational Linguistics (2018)
7. Iorio, A.D., Limpens, F., Peroni, S., Rotondi, A., Tsatsaronis, G., Achtsivassilis, J.: Investigating facets to characterise citations for scholars. In: Beltrán, A.G., Osborne, F., Peroni, S., Vahdati, S. (eds.) Semantics, Analytics, Visualization - 3rd International Workshop, SAVE-SD 2017, Perth, Australia, April 3, 2017, and 4th International Workshop, SAVE-SD 2018, Lyon, France, April 24, 2018, Revised Selected Papers. Lecture Notes in Computer Science, vol. 10959, pp. 150–160. Springer (2018)
8. Kacem, A., Flatt, J.W., Mayr, P.: Tracking self-citations in academic publishing. Scientometrics (2020)
9. Kayal, S., Afzal, Z., Tsatsaronis, G., Doornenbal, M.A., Katrenko, S., Gregory, M.: A framework to automatically extract funding information from text. In: Nicosia, G., Pardalos, P.M., Giuffrida, G., Umeton, R., Sciacca, V. (eds.) Machine Learning, Optimization, and Data Science - 4th International Conference, LOD 2018, Volterra, Italy, September 13-16, 2018, Revised Selected Papers. Lecture Notes in Computer Science, vol. 11331, pp. 317–328. Springer (2018)
10. Kayal, S., Groth, P., Tsatsaronis, G., Gregory, M.: Scientific topic attentionality: Influential and trending topics in science. In: Machine Learning, Optimization, and Data Science - 4th International Conference, LOD 2018, Volterra, Italy, September 13-16, 2018, Revised Selected Papers (2018)
11. Klavans, R., Boyack, K.W.: Research portfolio analysis and topic prominence. J. Informetrics **11**(4), 1158–1174 (2017)
12. Mayr, P., Frommholz, I., Cabanac, G.: Report on the 7th international workshop on bibliometric-enhanced information retrieval (BIR 2018). SIGIR Forum **52**(1), 135–139 (2018)
13. Mayr, P., Frommholz, I., Cabanac, G., Chandrasekaran, M.K., Jaidka, K., Kan, M., Wolfram, D.: Introduction to the special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL). Int. J. on Digital Libraries **19**(2-3), 107–111 (2018)
14. Mayr, P., Scharnhorst, A.: Scientometrics and information retrieval: weak-links revitalized. Scientometrics **102**(3), 2193–2199 (2015)
15. Min, C., Sun, J., Pei, L., Ding, Y.: Measuring delayed recognition for papers: Uneven weighted summation and total citations. J. Informetrics **10**(4), 1153–1165 (2016)
16. Mingers, J., Leydesdorff, L.: A review of theory and practice in scientometrics. Eur. J. Oper. Res. **246**(1), 1–19 (2015)
17. Otlet, P.: Traité de documentation : le livre sur le livre théorie et pratique / par Paul Otlet ; préf. de Robert Estivals, av.-pr. de André Canonne. Centre de lecture publique de la Communauté française de Belgique Ed. Mundaneum-Palais mondial (1934)
18. Panagopoulos, G., Tsatsaronis, G., Varlamis, I.: Detecting rising stars in dynamic collaborative networks. J. Informetrics **11**(1), 198–222 (2017)

19. Ravenscroft, J., Liakata, M., Clare, A., Duma, D.: Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. PLoS One **12**(3), e0173152 (2017)
20. Rouse, W.B., Lombardi, J.V., Craig, D.D.: Modeling research universities: Predicting probable futures of public vs. private and large vs. small research universities. Proceedings of the National Academy of Sciences **115**(50), 12582–12589 (2018)
21. Todeschini, R., Baccini, A.: Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research. Wiley-VCH (2016)
22. Wildgaard, L.E., Schneider, J.W., Larsen, B.: A review of the characteristics of 108 author-level bibliometric indicators. Scientometrics **101**(1), 125–158 (2014)