UNIVERSITY
INSTITUTE
OF LISBON

Beyond Citation Counts: Evaluating Publication Impact Through Journal-Level Citation Metrics

Guilherme Louro

Master in Computer Engineering

Supervisor:
PhD Fernando Manuel Marques Batista, Associate Professor
Iscte - Instituto Universitário de Lisboa

Supervisor:
PhD António Luís Morais Costa Silva Lopes, Assistant Professor,
Iscte - Instituto Universitário de Lisboa

September, 2025

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Beyond Citation Counts: Evaluating Publication Impact Through Journal-Level Citation Metrics

Guilherme Vaz Louro

Master in Computer Engineering

Supervisor:
PhD Fernando Manuel Marques Batista, Associate Professor
Iscte - Instituto Universitário de Lisboa

Supervisor:
PhD António Luís Morais Costa Silva Lopes, Assistant Professor,
Iscte - Instituto Universitário de Lisboa

September, 2025

Take Your Time

Acknowledgment

I would like to express my gratitude to the people who aided me in achieving this goal. Their help can't be understated when it came to being able to complete this work. Firstly, I'd like to thank my parents who were always available to help, no matter how minor the issue. Secondly, I'd like to thank my brother for his repeated words of wisdom and sage advice. Lastly, I'd like to thank my Marta for all of the care and attention she provided me through the difficulties I found during this phase in my life. I would also like to extend this gratitude to my colleagues and friends both at ISCTE and outside. With special consideration for my supervisors Fernando Batista and António Lopes, whose continued support and patience enabled me to be able to do this work. Their insights and motivations were invaluable in every step of this dissertation. Funding: This work was funded by FEDER–PT2020 partnership agreement under the scholarship reference Iscte_GDSI_02/2024.

Resumo

O impacto das publicações científicas é frequentemente avaliado através da contagem de citações, no entanto, essas métricas brutas podem não capturar adequadamente a verdadeira influência dos trabalhos citados. Para resolver isso, propomos uma metodologia que incorpora métricas de citação ao nível da revista para avaliar o impacto da publicação baseado no impacto das citações. Este estudo examina: "Citation Counts", "H-Index", "Journal Impact Factor" (JIF), "Eigenfactor Score", "Source Normalized Impact per Paper" (SNIP) e "SCImago Journal Rank" (SJR). Depois, foi construído um conjunto de dados que inclui artigos científicos selecionados, juntamente com os valores SJR, H-index e JIF dos seus artigos citantes. Este conjunto de dados foi apresentado numa tabela interativa e utilizado num inquérito a investigadores para avaliar qual das cinco métricas propostas reflete melhor as classificações destes especialistas. Das métricas propostas, duas baseiam-se no SJR, uma no H-Index e duas no JIF. A nossa investigação destacou o SJR como o indicador mais robusto, apresentando uma forte correlação com o JIF, uma das métricas mais amplamente utilizadas a nível das revistas científicas. Os resultados do questionário revelaram que a métrica de impacto de citações ajustada por JIF (Quadrática) foi a que melhor se alinhou com as classificações atribuídas pelos investigadores. No entanto, essa popularidade contrasta com a maior robustez da métrica de impacto de citações ajustada por SJR (Quadrática). Assim, este estudo estabelece uma base fundamental para investigações futuras sobre as métricas propostas, possibilitando a sua exploração e aperfeiçoamento.

Palavras Chave: *Impacto de Citações*; *Bibliometria*; *Métricas de Revista*; *Análise de Citações*; *SCImago Journal Rank (SJR)*

Abstract

The impact of scientific publications is often assessed through citation counts. However, these raw metrics may not adequately capture the true influence of cited works. To address this, we propose a methodology that incorporates journal-level citation metrics to evaluate publication impact. This study examines Citation Counts, H-Index, Journal Impact Factor (JIF), Eigenfactor Score, Source Normalized Impact per Paper (SNIP), and SCImago Journal Rank (SJR) to determine their effectiveness in reflecting the real impact of citations. Next, a dataset was constructed comprising selected scientific articles along with the SJR, H-Index, and JIF values of their citing papers. This dataset was presented in an interactive table and used in a researcher survey to assess which of the five proposed metrics best reflects these experts' rankings. Of these proposed metrics, two are based on SJR, one on the H-Index, and two on JIF. Our research found SJR as the most robust indicator, with a strong correlation to JIF, one of the most widely used journal-level metrics. Results from the survey indicated that the JIF-Adjusted Citation Impact proposed metric (Quadratic) most closely aligned with researchers' rankings. However, this popularity contrasts with the greater robustness of the SJR-Adjusted Citation Impact proposed metric (Quadratic), which is based on the Scimago Journal Rank. Consequently, this study establishes a foundational basis for future research on the proposed metrics, enabling further exploration and refinement.

KEYWORDS: *Citation Impact*; *Bibliometrics*; *Journal Metrics*; *Citation Analysis*; *SCImago Journal Rank (SJR)*

Contents

Acknowledgment iii

Resumo v

Abstract vii

List of Figures xi

List of Tables xiii

List of Acronyms xv

Chapter 1. Introduction 1
1.1. Background and motivation 1
1.2. Research Questions 2
1.3. Goals 3
1.4. Defining the Problem 3
1.5. Methodology 5
1.6. Dissertation Structure 5

Chapter 2. State of the Art 7
2.1. Citation Count Based Evaluations 8
2.2. Alternatives to Citation Counts 9
2.2.1. H-Index 9
2.2.2. Journal Impact Factor 11
2.2.3. Source Normalized Impact per Paper 12
2.3. Network-Based Alternatives 13
2.3.1. Eigenfactor Score 13
2.3.2. SCIMago Journal Rank 15
2.4. Correlation Between Metrics 16
2.5. Quality of Metrics 18
2.5.1. Artificial Inflation of Citation Counts 18
2.5.2. Manipulation of Authorship and Publishing 19
2.5.3. Summarised Comparative Analysis of the Metrics 20

Chapter 3. Proposed Approach 23
3.1. Early Experiments 23
3.2. First Steps of Automation of Data Gathering 24

Chapter 4. Development 29
4.1. The Metrics 32
4.2. Visualization of the Data 33

Chapter 5. Survey Answers 47

Chapter 6. Conclusion 53
6.1. Limitations and Future Work 54
6.2. Closing Remarks 55
6.3. Communication of Ethics Committee Approval 56
6.4. Communication of Research Dissemination 56

Bibliography 57

Appendix A. Researcher Survey 61
A.1. Citation Table 61
A.2. Researcher Survey 61

Appendix B. Ethics Committee Acknowledgment 67

List of Figures

Figure 1.1	System Pipeline	6
Figure 2.1	Example of an H-Index Calculation, taken from [**22**].	10
Figure 2.2	Journal Network with Prestige Transfer	14
Figure 2.3	Correlation Coefficients Meanings, taken from [**39**]	17
Figure 3.1	Data Pipeline Illustration	24
Figure 3.2	ISSN Selector	25
Figure 3.3	Visualization of the graph	27
Figure 3.4	SCImago Journal Rank (SJR) Heatmap Table	27
Figure 4.1	Table With SJR Values	29
Figure 4.2	Table with three columns per citation	30
Figure 4.3	Finalized Table with ISSN Values Corrected	31
Figure 4.4	Image of the table being sorted by SJR	35
Figure 4.5	Relationship Between Metrics	36
Figure 4.6	Rank Plot of SJR H-Index and Journal Impact Factor (JIF)	37
Figure 4.7	Rank Plot of SJR and JIF	38
Figure 4.8	SJR and JIF Behaviours	39
Figure 4.9	Behaviors With All Available Data	40
Figure 4.10	Metric Behaviours Normalized	41
Figure 4.11	Citation Data Correlations	41
Figure 4.12	Average Correlations	42
Figure 4.13	Pearson and Spearman Correlations of Proposed Metrics	43
Figure 5.1	Correlation between Answers and Metrics	49
Figure 5.2	RBO Similarities	50
Figure A.1	Page Presented to Researchers	62
Figure A.2	Proposed Metrics	62
Figure A.3	Section 1 of Survey	63
Figure A.4	Section 2 of Survey	64

Figure A.5 Section 3 of Survey 65
Figure A.6 Section 4 of Survey 65
Figure B.1 Ethics Committee Approval 71

List of Tables

Table 1.1	Table with SJR values collected on 02/07/2025	4
Table 1.2	Remaining Publications	4
Table 3.1	Journals with Contrasting H-Index Values	24
Table 4.1	Metric Rankings Across Publications	33
Table 4.2	Respondent's Publication Ranking	34
Table 5.1	Frequency of Metric Selection by Researchers	47
Table 5.2	Researcher and Quadratic and Cubic JIF Rankings	48

List of Acronyms

SJR: SCImago Journal Rank

JIF: Journal Impact Factor

SNIP: Source Normalized Impact per Paper

RBO: Rank Bias Overlap

CHAPTER 1

Introduction

The impact of a scientific publication is influenced by multiple factors, including the prestige of its authors, the reputation of the journal in which it is published, the number of citations it accrues, and the perceived influence of the citing sources. As a result, citation counts have emerged as a widely utilized metric for evaluating the impact of scholarly work. Numerous bibliometric indicators have been developed to assess impact at various levels, namely, the article, author, and journal levels. Prominent examples of such metrics include the H-index and the Journal Impact Factor [**22**, **41**].

As Agarwal et al. [**1**] observed that "the quantity of publications and their associated impact are commonly regarded as measures of a researcher's qualifications [...] The same principle applies to the reputation and standing of scientific journals." This assertion underscores a fundamental assumption in the application of bibliometric indicators for research evaluation: that a publication cited by high-impact journals may be presumed to hold greater influence than one cited by less prestigious sources.

To gauge the impact that citations have on scientific publications, we first needed to ensure the quality of the available metrics and automate the gathering of them. To guarantee this, we conducted a survey to ensure that the ordering of the publications made by researchers reflects the one made by the metrics.

1.1. Background and motivation

"Bibliometric methods estimate how much influence or impact a selected research article has on future research. It usually does this by counting the number of times the article is cited after it is published. The concept is that if a research article, called the source item, is cited in a future article, then it must have influenced the researchers who produced the future (downstream) article." This concept, originally articulated by Cooper in the work [**8**], forms a central foundation of the present study, whose primary objective is to calculate the impact mentioned in the quote. However, we don't want to do this just by using Citation Counts. We want to find and understand what is the best possible metric we can use to define impact.

To achieve this, journal-level metrics of the citing publications will be employed. As noted by SAGE, "Most researchers would prefer to publish an article in Nature than in this journal, but publication in this journal carries greater weight than in some other journals" [**14**]. In line with this rationale, the study operates under the premise that a publication from a highly prestigious journal is likely to have a higher impact than one from a less prestigious journal. This reasoning can similarly be applied to citations and

their influence on the publications they reference. Specifically, if a publication originates from a prestigious source, it is more likely to be impactful, furthermore, if it is frequently cited by other highly prestigious sources, it increases the likelihood that the publication itself was indeed impactful.

The indicators we'll use, aim to quantify the impact of scholarly output, and new metrics continue to emerge, each based on distinct methodologies. For the purposes of this study, we have chosen to examine the following metrics in relation to their utility in assessing the impact of individual publications: Citation Counts of the publication, the H-Index of the publishing journal, the Journal Impact Factor (JIF), the Eigenfactor Score, the SCImago Journal Rank (SJR), and the Source Normalized Impact per Paper (SNIP).

This selection aligned itself with the scope and objectives of the study, which was to develop a method for evaluating the impact of a publication by taking into account the impact of its citations. The approach relied on publicly available metrics, particularly those accessible via platforms such as Scopus, rather than proposing a novel or bespoke indicator created from the ground up.

With the target set of metrics established, it was necessary to gain a comprehensive understanding of each and assess their respective strengths and limitations. To achieve this, a review of these metrics was conducted, including documented examples of misuse and citation-related malpractices. This analysis informed the selection of metrics that appeared to offer a more balanced and reliable assessment of publication impact.

Therefore the work presented itself with the following steps: Understand which metric best reflects the impact of a publication based on its journal, gather these metrics, use them to calculate the impact of the cited publication and finally, have a metric that can calculate the true impact of a publication based on its citations.

1.2. Research Questions

In order to guide this investigation, a set of research questions was formulated to address the core challenges of evaluating publication impact through citation-based and journal-level metrics. These questions aim to clarify the effectiveness, robustness, and validity of established indicators when applied to the problem of assessing the true influence of a scientific publication. By examining both the quantitative properties of these metrics and their alignment with expert judgment, the study seeks to determine whether such measures can move beyond simple citation counts to provide a more comprehensive and trustworthy evaluation of scholarly impact.

RQ1: To what extent can journal-level bibliometric indicators (e.g., SJR, JIF, H-index, Eigenfactor, SNIP) provide an accurate representation of the scholarly impact of a publication, beyond the information conveyed by raw citation counts?

RQ2: How resilient are these indicators to manipulation strategies such as self-citation, citation cartels, and inflated authorship?

RQ3: How closely do metric-based assessments of impact align with expert researchers' qualitative judgments, and in what ways can bibliometric indicators be calibrated to best reflect expert evaluations and the true scholarly influence of a publication?

RQ1 addresses the fundamental task of understanding these metrics and ensuring that their intended representation aligns with the objectives of this study. RQ2 seeks to examine the ways in which these metrics may be undermined or manipulated. This is a critical consideration, as different metrics exhibit varying degrees of vulnerability to citation manipulation or so-called "citation gaming" practices. Finally, RQ3 is essential, as it aims to validate the extent to which the proposed metrics accurately reflect how researchers themselves assess the impact of a publication. This ensures not only that the evaluation of publications is both fair and robust, but also that it remains consistent with the perspectives and judgments of experts in the field.

1.3. Goals

The objective of the study was to provide a practical method for calculating the actual impact of citations, offering an alternative to relying solely on raw citation counts. The goal was to contribute a tool capable of more accurately reflecting the influence of a publication. To that end, data was collected on a sample of publications and their citing articles, incorporating metrics such as citation count and cumulative SJR values. This data set was then presented to a group of researchers, who were invited to assess the impact of the publications based on their own methodologies and perceptions. They were subsequently asked to indicate which metric most closely aligned with their assessments, thereby supporting an evaluation of the effectiveness of existing impact indicators.

An essential objective of this tool is to enhance, rather than replace, the value of a citation. While raw citation counts remain one of the primary indicators of an article's academic impact, the tool aims to supplement these counts with additional metrics that reflect the broader influence and significance of each citation. It is not our intention to diminish the role of citation frequency, but rather to introduce supplementary indicators that provide a more nuanced view of a citation's value. By incorporating these indicators, citations from high-prestige sources would be considered more impactful, while those from lower-prestige sources would be appropriately adjusted based on their metric-driven relevance. In essence, this approach fosters a dynamic transfer of prestige across citations, such that a citation with greater metric-based impact would contribute more substantially to the overall perceived impact of a publication, compared to one from a source of lesser prestige.

1.4. Defining the Problem

To demonstrate the problem we seek to tackle with our work, we provide a simplified table (Table 1.1) presenting the SJR values of citations. The listed publications were selected at random, and their Citation SJR values are arranged in descending order. The first column displays the publication title, including an embedded link to facilitate

manual verification. The second column indicates the number of citations received by each publication, followed by the SJR values of the top five citations.

TABLE 1.1. Table with SJR values collected on 02/07/2025

SJR values for each Cited Pub DOI						
Publication DOI	#Citations	C1	C2	C3	C4	C5
Publication 1	29	14.605	14.605	7.706	4.321	3.938
Publication 2	25	2.643	1.091	1.091	1.091	0.995
Publication 3	19	4.557	0.745	0.716	0.619	0.525
Publication 4	28	2.731	1.050	0.539	0.497	0.435
Publication 5	14	1.842	1.089	1.050	1.050	1.026

If asked to rank these publications based on the impact of their citations, the ordering would appear relatively straightforward at first glance: the first publication holds the highest number of citations as well as the highest SJR values. However, the ranking becomes more nuanced when comparing the fourth, third, and second publications. The third publication features a standout citation (C 1), while the fourth has the second-highest citation count. The second publication, though cited more frequently than the third but less than the fourth, exhibits the highest SJR values across all five citations. Such cases highlight the potential lack of coherence between rankings based solely on citation counts and those incorporating SJR values. This ambiguity underscores the need for a clearer understanding of how best to rank publications, which we will explore through our researcher survey.

To further illustrate these points, we include two additional citations in Table 1.2.

TABLE 1.2. Remaining Publications

Publication DOI	#Citations	C1	C2	C3	C4	C5
Publication 6	5	6.28	3.382	1.749	0.763	0.000
Publication 7	47	1.693	1.397	0.960	0.927	0.763

The inclusion of these two additional publications would significantly alter the ranking structure. Based on citation counts alone, Publication 6 would undoubtedly occupy the lowest position. However, an examination of the prestige of its citations reveals that its first two cites (C 1 & C 2) possess substantially higher prestige levels than the top five citations of most other publications in the sample, with the exception of Publication 1.

Our objective is not to disproportionately weight SJR values to the extent that Publication 6 surpasses all others in ranking except Publication 1. Rather, we seek to establish a sufficiently nuanced evaluation that enables Publication 6 to be meaningfully considered alongside more frequently cited but less prestigious publications.

1.5. Methodology

For the elaboration of this work, we built upon the system developed by Rodrigues [**37**], which provided an existing web scraping framework. This system was adapted to automatically collect data related to the journals citing a given publication. The pipeline was designed to begin with a user-supplied DOI, after which the web scraper retrieved all information regarding the citing publications, including the ISSN of the corresponding journals. Using ISCTE's CRIS, we then identified the year of each citation and obtained the associated SJR value for the citing journal.

Once all citation data had been gathered, it was necessary to ensure that the selected metrics aligned with how researchers themselves would rank publications, before proceeding with a system capable of fully implementing impact calculations. For this purpose, we initially compiled the SJR values into a reference table and subsequently expanded it to include the H-index and JIF. This approach allowed researchers to identify which of the metrics they considered most representative of the true impact of citations on a randomly selected set of scientific publications.

Presenting all available metrics for each citation, however, would have resulted in excessive information, complicating interpretation. To address this, and to balance citation counts with journal-level metrics, we proposed five new composite metrics that integrate the available indicators for each citation. Furthermore, we introduced a formula that ensures publications with high citation counts but relatively low indicator values can still be regarded as impactful, while also allowing publications with fewer citations but from highly prestigious journals to reflect their influence appropriately. This approach provides a more balanced interpretation of the impact of citations on scientific publications.

1.6. Dissertation Structure

Following this introductory chapter, the study is structured as follows. Chapter 2 provides a literature review of existing journal-level metrics and discusses the ways in which these can be manipulated. It also examines the relationships and correlations between selected metrics.

Chapter 3 outlines the initial experiments and describes the development of the plan to create a system capable of presenting metrics to researchers. This plan involved the design of an interactive, sortable table displaying the metrics of the journals of cited publications alongside their citation indicators, with the aim of determining which ordering best reflected the rankings established by researchers.

Chapter 4 explains how these plans were implemented and discusses the challenges encountered throughout the process. In addition, this chapter details the proposed metrics presented to researchers in the table, as well as the correlations identified among the various indicators and metrics.

Chapters 5 and 6 present the survey responses, correlating the rankings provided by researchers with those derived from the proposed metrics. These chapters also draw the

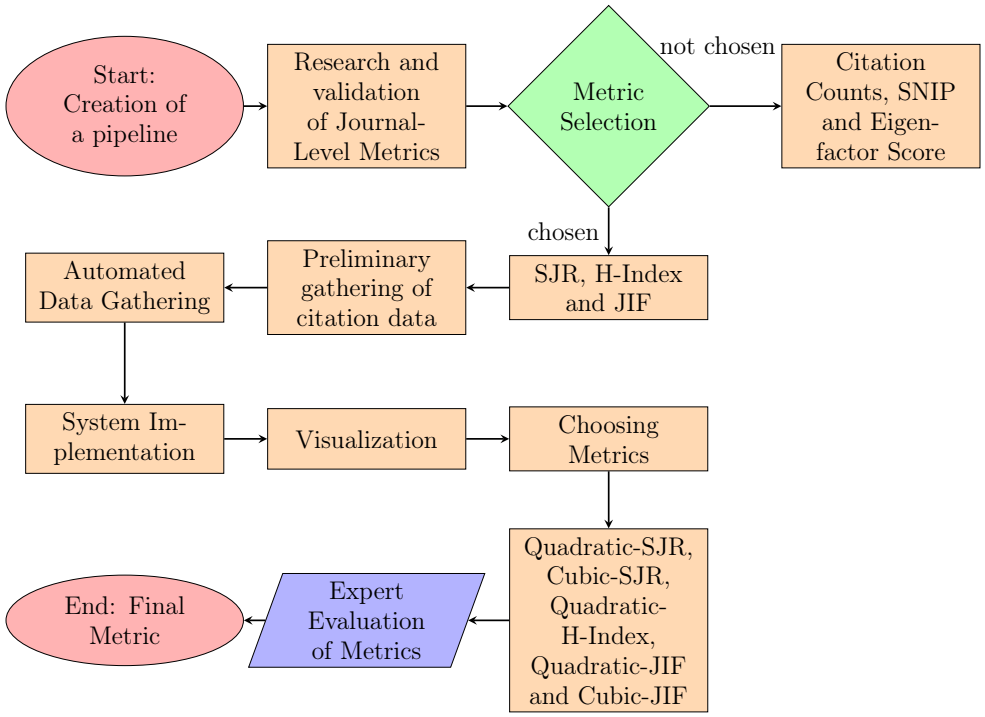


FIGURE 1.1. System Pipeline

main conclusions of this work, discuss its limitations, and highlight directions for future research.

Following these chapters is the bibliography containing all referenced works is presented, followed by the Appendix A, Researcher Survey, which details the material provided to the researchers in our survey.

CHAPTER 2

State of the Art

Citation metrics and bibliometrics about the impact and influence of an article are centered around the concept already established in Section 1.1, that when a publication "is cited in a future article, then it must have influenced the researchers who produced the future (downstream) article. Being cited by another researcher indicates that the source researcher is having an impact on the science: The research product is being used by others to create even more information" [**8**]. This conceptual foundation underlies the use of metrics designed to quantify the influence of a publication. Furthermore, we wished to give more importance to the impact of a publication without solely relying on the amount of citations to attempt to circumvent the issue that is exposed in [**45**]. While a publication with many citations is definitely something we wish to prioritize, we also want to make sure to give priority to citations of high impact which circles back to the aforementioned quote from [**8**].

The work presented in [**40**] served as an important point of comparison for the present study. That work aimed to calculate the impact of an author by considering the impact of the authors citing them. In a similar manner, the objective of this thesis is to evaluate the impact of a publication rather than that of an author. Although the approaches are not identical, the study offers several relevant parallels, particularly in its attempt to design a metric resistant to citation manipulation. The proposed metric in [**40**] was frequently contrasted with the H-index, thereby underscoring the latter's shortcomings and demonstrating its susceptibility to citation gaming strategies. By contrast, the p-index sought to mitigate these vulnerabilities through the incorporation of a PageRank-based component.

A key distinction between this prior work and our own lies in the nature of the metric itself: the p-index constitutes a fully bespoke metric, whereas our approach seeks to leverage already established and validated metrics to assess the impact of a publication. Another notable similarity, however, is the shared emphasis on evaluating the influence that a citation has on a publication, reflecting the transfer of prestige that occurs when one article cites another, a concept that closely aligns with the PageRank-based formulation underlying the p-index.

Since we wish to diverge from the p-index and use existing journal-level metrics, we first needed to identify which metrics are most appropriate for this and like with the p-index, understand how robust they are in comparison to each other against manipulation and gaming.

Our work centered on the use of a metric that is able to quantify the impact that citations have on scientific publications through journal level indicators of impact. To re-iterate, the indicators that we chose to focus on for our work were: raw citation counts, the H-Index, the JIF, SNIP, Eigenfactor Score and SJR. These indicators were chosen because they are widely available, popular, attempt to, and are used to calculate the impact of journals. Given our goal can also be explained as the transfer of prestige through citations, both Eigenfactor Score and SJR were put into focus. Another obvious metric that could be picked is Scopus' CiteScore, but due to that metric's similarity to JIF, it was set aside.

Before exploring the metrics mentioned and understanding, it is important to thoroughly explain raw citation counts. All of the metrics mentioned are, at heart, citation-based journal level metrics even if their calculations may take different forms [**33**, **38**].

2.1. Citation Count Based Evaluations

According to Cooper [**8**], citation counts represent the foundational metric of bibliometric analysis, defined simply as the number of times a given source has been cited. As he notes, "Bibliometrics does not exist without citation counts."

As mentioned above by [**8**], in bibliometrics, a citation is understood as a transfer of knowledge, such that impact is established when one article references another. For this reason, citation counts are widely employed as indicators of an article's impact: the more frequently an article is cited, the greater its presumed influence on subsequent publications [**46**].

However, simple citation-based evaluation constitutes a rather unstable foundation. As will be examined in greater detail in the subsequent chapters, citation counts can often provide a misleading estimation of the actual impact of an article or journal. Several factors contribute to this, one of which is that citation counts, once accumulated, never decrease making it so there is no factor that limits anomalous citation behaviors. Consequently, older journals may continue to benefit from the prestige of their past publications, whereas newer authors and journals, which have not yet accumulated a significant citation record, may be overlooked regardless of the relevance or impact of their contributions [**1**, **7**, **17**].

Having established that Citation Counts are both flawed as a measure of scientific impact and one of the most popular indicators of such impact it is important to also see some of the advantages and disadvantages of using Citation Counts.

The main pros of Citation Counts rest in great part in their ease of access, simplicity and popularity. Firstly, the ease of access: Citation Counts are just a tally of the number of citations made from one journal, author or article to another. Indexing databases such as Scopus, Web of Science, Google Scholar or OpenAlex make the citations made to an article available. This number may vary from one database to another, but it is able to give an idea of how many citations one article has received [**8**]. Secondly, the simplicity and popularity, citations and the influence that one article has on another article is easy

to interpret, not only that but Citation Counts have been a staple way to gauge the impact of a publication for decades. This results in a well established and simple way to understand the impact, not the quality of a publication [**38**, **42**].

As previously noted, citation counts never decrease in value, which leads to legacy journals being perceived as more impactful than newer journals, even when the latter may publish work of greater significance to the scientific field.

Another limitation of citation counts is the uniform value assigned to each citation: a reference from a highly influential journal such as *Nature* is weighted the same as a citation from a less influential source. This uniformity creates opportunities for common citation gaming practices, such as excessive self-citation [**23**]. In such cases, a journal may publish articles that disproportionately cite its own works in order to artificially inflate its citation counts. Unlike legitimate citations, which reflect the transfer of knowledge between articles, these practices represent attempts to manipulate metric values.

The issue of uniform citation value extends beyond citation counts and, to some degree, affects network-based metrics that will be discussed in subsequent sections. It is precisely this uniform valuation that has given rise to a variety of citation gaming methods.

Examples of such practices, in addition to excessive self-citation, include citation padding, where authors insert superfluous references to artificially increase citation numbers; citation cartels—where journals collude to systematically cite each other; and excessive multi-authorship, since each author receives equal recognition from a citation, producing an incentive to join works likely to accrue numerous citations, even without substantial contributions. As citation counts form the foundation of many widely used bibliometric indicators, these limitations and manipulative practices inevitably permeate into other derived metrics as well [**10**, **11**, **24**].

2.2. Alternatives to Citation Counts

Knowing that raw citation counts give a flawed idea of the impact of a publication it is necessary to find a more robust metric, one that can more accurately in concept find the true impact that a publication had via the journal level metrics of the citations. Another thing to keep in mind is our true objective with these metrics. Our main goal was to use a metric be it altered or unaltered to quantify the impact that a publication has based on the impact of its citations. In other words we want to find the metric that best reflects the transfer of prestige that occurs when one publication cites another.

2.2.1. H-Index

Introduced by Hirsch the H-index [**22**] is "an easily computable index, h, which gives an estimate of the importance, significance, and broad impact of a scientist's cumulative research contributions." It is calculated by determining the number h of a journal's articles that have received at least h citations. Figure 2.1 illustrates the process.

The H-Index constitutes one of the most important metrics examined in this work, primarily due to its widespread popularity and the significant body of literature that both

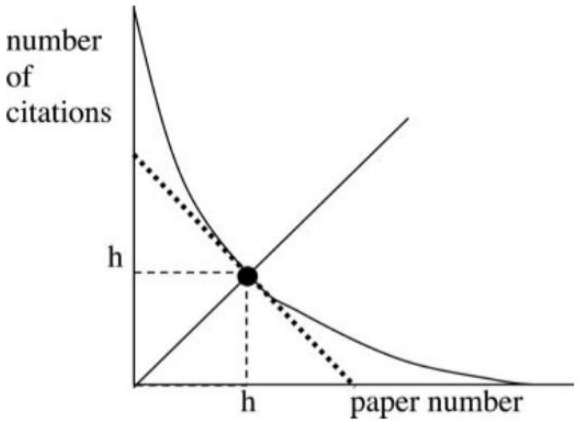


FIGURE 2.1. Example of an H-Index Calculation, taken from [**22**].

supports and criticizes its use. Moreover, it has served as the basis for the development of numerous derivative indices, such as the g-index, AR-index, and H10-index. Despite these subsequent efforts to refine or improve upon the H-Index, their limited adoption has meant that the H-Index remains the most influential and widely recognized among them [**9**, **25**].

The H-Index offers a more nuanced perspective on publication impact than raw citation counts alone, as it reflects not only the volume of work produced but also the extent to which that body of work has been influential. In theory, this allows the index to highlight journals and authors who are both active and impactful. Furthermore, due to the nature of its calculation, the H-Index is time-insensitive, meaning that its value cannot decrease once established.

Nevertheless, the index is fundamentally determined by the number of citations, and as such, it remains vulnerable to practices such as excessive self-citation. Another limitation of the H-Index arises in its treatment of journals that publish impactful articles infrequently: these journals may receive lower H-Index values than journals producing larger numbers of publications with varying levels of impact. This issue will be further explored later in this work. A similar line of reasoning applies to the comparison between older and newer journals. Newer journals, regardless of the quality or impact of their publications, face structural disadvantages relative to older journals, whose accumulated body of work tends to secure a higher H-Index. This is reinforced by the fact that the H-Index does not decrease over time. As a result, larger and more established journals may appear more impactful even in cases where they no longer produce highly influential work [**9**, **49**].

A final shortcoming of the H-Index is its tendency to undervalue highly cited individual articles. No matter how influential a single article, or a small group of articles, may be, their contribution becomes negligible if not accompanied by a sufficient number of other publications [**9**]. At first glance, this might seem contradictory to our argument that such a characteristic constitutes a limitation of the H-Index within the scope of this research. Our aim is to provide as balanced an assessment as possible of a publication's impact

based on its citations. From this perspective, the insensitivity of the H-Index to outlier publications with exceptionally high citation counts could appear beneficial. However, in practice, this runs counter to our objectives. The intention of this work is not to diminish the value of highly popular publications; rather, it is to highlight and promote such influence where it exists, while guarding against instances in which apparent popularity is achieved through questionable or manipulative practices. In this sense, the objective is not to devalue citations themselves, but instead to construct a framework in which both popularity and prestige are weighted fairly, ensuring that journals with genuine scholarly impact are distinguished from those whose influence may be inflated through less reputable means.

Taken together, these factors contribute to a predominantly negative assessment of the H-Index within the framework of this research. Many of the limitations associated with raw citation counts continue to manifest within the H-Index, thereby allowing their effects to persist and potentially distort the perceived impact of publications. Nevertheless, the H-Index remains one of the most widely used and recognized metrics in the academic community. For this reason, although it cannot serve as the principal means of measuring publication impact in this work, it also cannot be disregarded and must be considered alongside other indicators.

2.2.2. Journal Impact Factor

One of the two metrics we chose to investigate that is publicly accessible via Clarivate's journal Citation Reports, the IF was originally designed to facilitate comparisons between journals regardless of their size. As described by Cooper in [**38**], the IF is calculated by dividing the number of citations in the current year to articles published in the previous two years by the total number of citable items published during that same period. As can be seen in page 5 Figure 2 of [**13**].

$$\mathbf{IF}_y = \frac{\text{Citations}_y}{\text{ArticlesPublished}_{y-1} + \text{ArticlesPublished}_{y-2}}$$

Another highly relevant metric examined in this study is the JIF. Similar to the H-Index, it enjoys widespread popularity and is frequently employed by institutions as a basis for decisions concerning journals and their articles. Historically, the JIF was originally introduced as a tool to assist libraries in determining which journals to subscribe to. Unlike the H-Index, the JIF provides a different perspective on scholarly impact, focusing on short-term influence by assessing the immediate citation performance of a journal, thereby offering insight into current impact rather than the longer-term perspective afforded by the H-Index [**27**].

Despite its widespread recognition, the JIF is not without limitations. The standard calculation does not account for excessive self-citations, which can artificially inflate values. Clarivate has attempted to address this issue by offering an alternative version of the JIF that excludes self-citations entirely. According to [**25**], "The journal self-citation means the case where a paper published in the journal J is cited in the same journal."

However, this adjustment is also problematic, as self-citations are an inherent and often necessary part of scientific communication [**49**]. Another variation, offered by Thomson Reuters, employs a five-year citation window instead of two years, thereby accommodating journals whose publications require more time to reach their citation potential. Nonetheless, the two-year version remains the most widely used and recognized in academic evaluation [**26**].

A defining feature of the JIF is the direct relationship between citation counts and the value of the metric: the higher the number of citations, the higher the JIF. While versions that reduce the impact of self-citations do exist, the metric remains vulnerable to citation manipulation strategies aimed at artificially inflating citation counts. One example is "citation stacking," described in [**21**] as "the behavior of a network of journals, which work apparently together to cite each other and thus raise their impact factors."

Because the JIF is based on an average, it is particularly susceptible to distortion by a small number of highly influential articles. In such cases, a journal may appear highly prestigious due to a handful of widely cited works, while the majority of its publications exert little to no influence [**2**, **25**, **26**, **41**]. Although this phenomenon can be viewed as a limitation, it may also be interpreted as a natural reflection of the uneven distribution of impact across publications in academia, rather than as an inherent flaw in the metric.

In conclusion, while the Journal Impact Factor is not free from methodological weaknesses and remains exposed to certain citation malpractices, it nonetheless provides a meaningful and broadly comprehensible measure of journal impact. Its relatively straightforward calculation and widespread recognition render it particularly relevant for the purposes of this work.

2.2.3. Source Normalized Impact per Paper

Available through Scopus, the SNIP adjusts for variations in citation practices across different academic fields. As described by Moed [**32**], it "measures a journal's contextual citation impact," accounting for factors such as citation frequency, citation maturity, and database coverage. The SNIP is calculated as the ratio between a journal's Raw Impact per Paper (the average number of citations to articles that are one to three years old) and the Database Citation Potential (the average number of one- to three-year-old references in citing articles across the database).

The SNIP thus provides an alternative interpretation of publication impact by incorporating field-specific citation behavior. This enables more meaningful comparisons across disciplines and research areas, offering a field-normalized perspective on journal performance.

Nonetheless, the metric has been subject to criticism. One common critique relates to the complexity and lack of transparency in its calculation. While such concerns are valid, a more fundamental limitation of the SNIP in the context of this research has been identified in [**7**], where the authors state: "Moreover, if one considers that it is appropriate to weight citations on the basis of the status of the citing journal, SJR is

a better indicator than SNIP. On the other hand, if one considers that journal impact and topicality or citation potential are two distinct concepts that should be measured and assessed separately, SNIP and its components, especially raw impact per paper and citation potential, are more informative."

This observation highlights the intrinsic misalignment of the SNIP with the objectives of this study. Since the aims of this work involve assessing impact through both popularity and prestige, SNIP does not provide the most suitable framework. Consequently, the focus will shift to the two remaining network-based metrics, which align more directly with the goals of this research.

2.3. Network-Based Alternatives

"The idea of these indicators is that having a paper cited in a very high-quality or prestigious journal such as Nature or Science is worth more than a citation in an obscure journal" [**30**]. These metrics are all based on the Pagerank algorithm so it makes sense to introduce it conceptually prior to explaining the metrics:

The PageRank algorithm can be widely recognized as the mechanism originally employed by Google to rank web pages. This process can be conceptualized by abstracting web pages as nodes and the hyperlinks connecting them as edges. Within this abstraction, the broader applicability of the algorithm, as outlined in [**34**], becomes evident. One such application concerns scientific publications, where journals can be modelled as nodes and citations between them as edges. This perspective enables the ranking of journals through a network-based approach such as PageRank. Regarding its computation, [**6**], describe PageRank as "an iterative algorithm which propagates prestige values from one web page to another and converges to a solution." In practice, this corresponds to an algorithm that distributes prestige across the network of nodes. PageRank also addresses challenges such as sink or dangling nodes where a node receives prestige but lacks outbound edges to redistribute it. This issue is mitigated through the introduction of a damping factor, typically, which ensures that with probability d prestige follows outgoing edges, and with probability (1 – d) it "teleports" to a random node. This mechanism, often described as random walks, not only prevents prestige from accumulating in dangling nodes but also reflects how users may arrive at a page or, in the scholarly context, a journal directly through search engines or indexing databases [**28**, **36**].

2.3.1. Eigenfactor Score

Based on the PageRank algorithm, the Eigenfactor Score employs citation networks to evaluate a journal's influence by assigning greater weight to citations originating from more influential journals. As Bergstrom explains in [**3**], "We measure the importance of a citation by the influence of the citing journal divided by the total number of citations appearing in that journal."

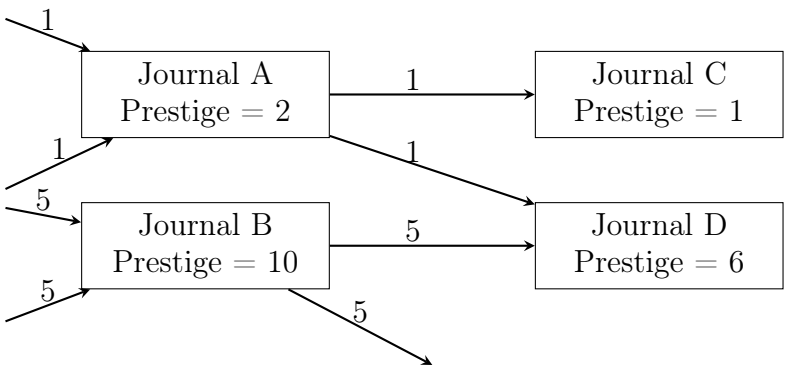


FIGURE 2.2. Journal Network with Prestige Transfer

A useful distinction between the Eigenfactor Score and the Journal Impact Factor is highlighted in [**13**]: "The Eigenfactor measures the influence of the journal, but not the influence per article as the IF does."

The principal advantages of the Eigenfactor Score arise from the nature of its calculation [**12**]. Since it relies on a network-based approach, the prestige of the journals providing citations is the main determinant of impact, rather than the sheer number of citations. While not entirely immune, the Eigenfactor Score reduces the influence of artificially inflated citation counts compared to the Journal Impact Factor.

Another mechanism through which the Eigenfactor Score mitigates citation manipulation is its complete exclusion of self-citations [**35**, **48**]. However, as discussed previously in relation to the Journal Impact Factor, this approach also has drawbacks, since many self-citations occur as a natural and legitimate part of scientific research rather than solely as attempts to increase indicator values.

A frequently raised critique of the Eigenfactor Score is that a larger volume of publications disproportionately improves a journal's score [**30**, **38**, **49**]. This occurs because the calculation does not include a denominator; that is, the Eigenfactor Score is never normalized by the number of publications. As [**25**] explains, "the sum of all Eigenfactor scores for all journals in the database is equal to 100," and, "Since this quantity is not normalized by the total number of papers published in a given journal, it tends to be larger for journals publishing larger numbers of papers."

This issue can be addressed by the Article Influence Score, which normalizes the Eigenfactor Score by the number of publications produced by a journal over the previous five years. As described in [**3**, **31**], "The Article Influence is scaled with respect to a mean of 1.00."

Another strength of the Eigenfactor Score is its use of a five-year citation window, which provides articles with a longer timeframe to accrue citations. This feature offers a potentially more accurate representation of impact for journals in fields where citation accumulation is slower and more procedural, such as mathematics and ecology. By contrast, the two-year citation window used in the Journal Impact Factor can be limiting in such disciplines [**48**].

Overall, the Eigenfactor Score offers a more compelling perspective on the assessment of impact within the scope of this research. Its capacity to capture the transfer of influence within citation networks aligns closely with the objectives of this study. Nonetheless, one additional metric remains to be considered, which seeks to achieve a similar goal through a different approach.

2.3.2. SCIMago Journal Rank

Also derived from the PageRank algorithm and available via Scopus, the SJR measures journal prestige based on the influence of the citing journals. As explained in [**7**], "when a journal A is cited, say, 100 times by the most highly ranked journals in the field, it receives more prestige than a journal B that also receives 100 citations, but from less prestigious periodicals."

The SJR metric is calculated in two phases. In comparison to the Eigenfactor Score, the second phase of the computation brings SJR closer to the Article Influence Score, due to its normalization, which makes the SJR "a size-independent metric, [...] which can be used to compare journals" [**15**]. Another distinction lies in the citation window: whereas the Eigenfactor Score relies on a five-year window and the Journal Impact Factor on two years, SJR adopts a three-year citation window. As demonstrated in [**15**], this was identified as the shortest time frame that still "embraces citation peaks of all the subject areas in Scopus."

The version of SJR currently available through Scopus is SJR2, which introduced several methodological improvements. Specifically, the value of prestige transferred is now determined using the cosine of the co-citation profile between two journals, ensuring that citations from more thematically distant journals carry less weight. Moreover, "The transfer of prestige to another journal or to itself is limited to a maximum of 50% of the prestige of the journal source, and a maximum of 10% per citation." Thus, a single citation can account for no more than 10% of the prestige of the citing journal, and the total amount of transferred prestige cannot exceed half of the citing journal's prestige. This approach effectively constrains excessive cross-citation practices, such as citation farms. Additionally, prestige from dangling nodes is now redistributed proportionally according to what journals receive from citing journals. The second phase of the calculation normalizes exclusively to citable documents, and short surveys are included as cited items [**18**].

The self-citation cap, rather than a full exclusion, provides a more balanced approach that better reflects the realities of scientific publishing. "Next, in order to prevent excessive journal self-citation, the number of references that a journal may direct to itself is limited to a maximum 33% of its total references" [**15**]. In theory, this limits excessive self-citation, while maintaining some given their natural occurrence in academia.

The calculation of SJR is outlined in [**18**], along with guidance for interpreting its values. Since the metric is normalized, an SJR of 1 corresponds to the mean level of impact across journals, an SJR of 2 indicates a journal twice as prestigious as the mean, and an SJR of 4 reflects four times the average prestige. However, this description is the

only publicly available information on the SJR's computation. As noted in [**29**], while the source code used to compute the Eigenfactor Score is openly accessible, the SJR methodology is only documented through formulae, raising concerns regarding reproducibility and transparency.

Another important distinction between the SJR and the Eigenfactor Score is the data source on which each metric is based. SJR relies on Scopus and is disseminated through the openly accessible SCImago website, whereas the Eigenfactor Score draws on data from Clarivate's Journal Citation Reports, which requires a subscription.

Although SJR has also been criticized, for example regarding the somewhat arbitrary nature of certain elements of its calculation, it is nonetheless subject to fewer methodological weaknesses than other indicators. As emphasized in [**7**], "if one considers that it is appropriate to weight citations on the basis of the status of the citing journal, SJR is a better indicator than SNIP." For these reasons, and due to the generally robust qualities of the metric, SJR is regarded in this work as a particularly promising indicator of the impact of citations on scientific publications. Its similarity to the Eigenfactor Score and Article Influence Score further strengthens its relevance, while its refinements position it as the most compelling of the three.

2.4. Correlation Between Metrics

While most of the metrics analyzed offer distinct interpretations of what constitutes impact and how it should be measured, it is equally important to understand how these indicators relate to one another in practice. To address this, several studies have investigated the degree of correlation between the principal bibliometric indicators.

The most common approaches employed in these studies are the Pearson and Spearman correlation coefficients. These statistical measures evaluate the degree of association between two continuous variables, making them naturally applicable to pairs of citation-based metrics. The Pearson coefficient assesses linear relationships, whereas the Spearman coefficient captures monotonic, potentially non-linear relationships. Figure 2 in [**39**] illustrates the practical differences between these two measures with two graphs.

In both cases, the interpretation of the coefficients is similar: values range from -1 to 1, with negative values indicating inverse associations and positive values indicating direct associations. Conventionally, correlations in the range of 0.9 to 1.0 are considered very strong [**39**, **43**].

Empirical evidence suggests consistently high correlations between SJR and modified versions of the JIF. Specifically, early analyses reported strong correlations between SJR1 and a three-year citation-window JIF constructed for comparative purposes. Similar results were obtained with SJR2, again demonstrating strong correlations with a three-year JIF. In these studies, Pearson correlations were generally lower than their Spearman counterparts, though both confirmed strong positive associations. Reported values for the Pearson coefficient were 0.882 globally, 0.910 across 27 subject areas, and 0.873 across 295 specific subject categories. The corresponding Spearman coefficients were 0.944, 0.944,

Table. Example of a Conventional Approach to Interpreting a Correlation Coefficient

Absolute Magnitude of the Observed Correlation Coefficient	Interpretation
0.00–0.10	Negligible correlation
0.10–0.39	Weak correlation
0.40–0.69	Moderate correlation
0.70–0.89	Strong correlation
0.90–1.00	Very strong correlation

FIGURE 2.3. Correlation Coefficients Meanings, taken from [**39**]

and 0.917, respectively, further confirming the strength of the relationship between these indicators [**15**, **18**].

Additional evidence from the field of neurosurgical journals reinforces this pattern, with high Spearman correlations observed between SJR and JIF, JIF excluding self-citations, and the H-index. Although the correlation between H-index and JIF was slightly lower, it still represented a strong positive association [**39**]. Eigenfactor Score also demonstrated positive correlations with other metrics, though its coefficients were generally lower in magnitude [**49**].

A broader comparative analysis presented in [**44**] highlights an especially consistent and strong correlation between Eigenfactor Score and raw citation counts. This finding corroborates one of the main criticisms of the Eigenfactor: its lack of normalization relative to the number of published articles. By contrast, SJR was shown to correlate strongly or very strongly with JIF across four of six subject areas, and also displayed high correlations with the Article Influence Score. However, the authors did not specify which correlation coefficient was used, leaving some ambiguity in the interpretation of these results.

Overall, SJR has been shown to correlate highly with JIF, despite their conceptual differences in framing journal impact. Both metrics generally exhibit similar trends, and according to [**18**], both follow a logarithmic distribution when normalized. Although their absolute values differ in magnitude, the two indicators display somewhat parallel behaviors across a wide range of academic fields.

Having established and understood these findings, we can now address Research Question 1:

To what extent can journal-level bibliometric indicators (e.g., SJR, JIF, H-index, Eigenfactor, SNIP) provide an accurate representation of the scholarly impact of a publication, beyond the information conveyed by raw citation counts?

We now recognize the degree to which these metrics improve upon raw citation counts, while also acknowledging their respective limitations. For example, the H-index and JIF, although widely used, are both popular yet fundamentally flawed indicators. The SNIP metric, while valuable in certain contexts, diverges from the objectives of this work. The

Eigenfactor Score addresses some of the limitations of the JIF by employing a citation network–based methodology; however, its lack of normalization makes it vulnerable to manipulation through excessive authorship and publication volume. By contrast, the SJR emerges as the most robust metric, as it explicitly accounts for the transfer of prestige across citations, thereby aligning closely with the goals of this study and has many forms of accounting for possible malpractice.

Nevertheless, in order to fully address RQ1, it is essential to investigate in greater depth how these metrics may be manipulated. Such an examination not only clarifies their vulnerabilities but also directly contributes to answering RQ2.

2.5. Quality of Metrics

"The universal academic mantra, to the point of cliché, is "publish or perish." [...] The pressure behind the publication drive is twofold – the enhancement of one's curriculum vitae for the purposes of promotion and for the sake of academic and professional repute, and for plausibility when applying for grants/funding" [**16**]. This pressure is often at the heart of the decision making that leads researchers to use citation manipulation.

As articulated by Biagioli in [**5**], "Goodhart's law (named after the British economist who may have been the first to announce it) states that when a feature of the economy is picked as an indicator of the economy, then it inexorably ceases to function as that indicator because people start to game it." This principle applies directly to citation-based metrics used in the evaluation of scholarly impact. Although such metrics aim to quantify academic influence, they are not immune to manipulation. journals, in particular, may adopt practices specifically designed to artificially inflate their performance on these indicators [**24**].

In the following section, we will explore some of the methods by which citation counts can be artificially inflated.

2.5.1. Artificial Inflation of Citation Counts

Metrics such as Citation Counts, the H-Index, and the JIF are fundamentally dependent on the number of citations a publication receives. These indicators do not account for the origin or context of citations, thereby rendering them susceptible to malpractice, including citation rings, coercive citation, citation padding, and excessive self-citation [**11**, **21**, **24**]. Citation rings and coercive citation often involve collusive arrangements among journals that disproportionately cite each other, or explicit citation exchange agreements intended to boost their respective metrics [**10**, **11**]. Similarly, citation padding and excessive self-citation artificially elevate indicator values through internal referencing and repetitive citations within networks of associated journals. As Heneberg in [**21**] explains:

"As 'network self-cites' we understood any citations received by the analyzed journal from other journals within the 'network'. As a 'skewed pattern of cites by self-cites' we understood the citation ratios, which were strongly affected, i.e., increased, by the presence of journal self-cites. [...] As 'citation stacking' we understood the behaviour of

a network of journals, which work apparently together to cite each other and thus raise their impact factors."

These practices directly and substantially influence citation-dependent metrics. The Journal Impact Factor and the H-Index are particularly vulnerable, as both rely on proportional relationships between the number of citations received and either the number of publications or threshold citation counts [**4**, **41**].

By contrast, alternative metrics such as SNIP, Eigenfactor Score, and SJR adopt methodologies that mitigate these vulnerabilities. SNIP adjusts for field-specific citation behavior, scaling the weight of each citation according to disciplinary citation norms. This contextualization reduces the influence of artificially inflated citation practices. Eigenfactor Score and SJR instead apply network-based approaches, weighting citations by the prestige of the citing source. In these systems, not all citations are treated equally, reducing the effectiveness of manipulative practices [**38**].

Given these characteristics, Citation Counts and the H-Index appear particularly susceptible to gaming, since they assign equal weight to all citations and are thereby easily inflated by sheer volume. The Journal Impact Factor introduces a corrective mechanism by relating citations to recent publication volume. However, its proportional relationship with raw citation counts similarly constrains its robustness and fairness as a reliable measure of scholarly impact.

Consequently, attention should be directed toward metrics that account for citation quality and contextual impact. SNIP is notable for emphasizing cross-disciplinary influence and assessing impact across different fields; however, this perspective lies outside the scope of the present study. Greater focus is therefore placed on metrics that capture the transfer of impact through citations, such as SJR. The Eigenfactor Score is also recognized for its robustness, due to its foundation in the PageRank algorithm and its emphasis on citation influence rather than sheer quantity. Nonetheless, its lack of normalization renders it susceptible to certain manipulation tactics. Such as the ones highlighted in the following section.

2.5.2. Manipulation of Authorship and Publishing

Another set of manipulative practices concerns the authoring and publishing side of scholarly production, with strong implications for the H-Index and Eigenfactor Score. For the former, the key determinant is not only the number of citations per publication but also the total number of publications produced [**22**]. For the latter as was discussed earlier a more productive journal usually nets a higher Eigenfactor Score since there is not normalization present that ponders the amount of citation with the number of publications, only the impact of the citation matters for the calculation [**25**, **30**, **38**, **49**].

One notable tactic is "salami slicing," whereby a research output is intentionally divided into multiple smaller publications. This strategy maximizes citation opportunities, since one article naturally leads to another, and external authors may be compelled to cite several related papers in order to cover a single topic comprehensively [**24**]. While

such practices can substantially inflate a journal's H-Index, they often do so without the goal of genuinely advancing the field. This practice, naturally, can also substantially influence the Eigenfactor Score, as an author who publishes a work divided into multiple smaller segments will generate citations to these numerous sliced articles, thereby accruing a greater number of citations than would be warranted by the original research output.

The Journal Impact Factor, in contrast, is less directly affected, as it balances the number of publications against the number of citations received. Thus, publishing many low-citation papers can be detrimental to the JIF, since it reduces the average citations per publication.

Other manipulative practices include inflated multi-authorship and "gift" authorship. The latter extends also to citations. In both cases, the central assumption is that the mere presence of an author's name guarantees credit in citation-based evaluations, regardless of the individual's actual contribution. This allows journals to inflate their indicators without proportionate investment in research quality or output [**11**]. Such practices can distort all metrics, since none are designed to account for differentiated contributions among authors [**4**]. However, JIF and SJR both mitigate this malpractice somewhat since they both normalize their metric values with the number of publications in their time windows. Unlike Eigenfactor or H-index, for example.

To sum up, ultimately, SJR was selected as the primary metric for this study due to its network-based methodology, emphasizing the transfer of prestige across citations. Which aligns closely with the conceptual framework of our research and offers inherent resistance to citation manipulation. Unlike raw citation counts or the H-Index, SJR incorporates prestige transfer, imposes a cap on self-citations, and discounts instances where self-citations exceed one-third of a journal's total [**3**]. Furthermore, SJR is updated annually and is readily accessible via Scopus, making it both methodologically sound and practically suitable for our objectives.

2.5.3. Summarised Comparative Analysis of the Metrics

- **Citation Counts [42]**
  Description: Measures article popularity by number of citations.
  Pros: Widely available; Intuitive.
  Cons: Never lowers in value; Easily manipulated.
- **H-Index [9, 22, 38]**
  Description: Balances productivity with citation counts.
  Pros: More robust than citation counts.
  Cons: Never lowers in value; Easily manipulated.
- **JIF [13, 38, 41]**
  Description: Average citations to articles from previous two years.
  Pros: Widely used and popular; Captures dissemination balanced with productivity.

Cons: Easily gamed (e.g., excessive self-citations); Two-year citation window is limiting.

- **Eigenfactor Score [31, 38, 48, 49]**
  Description: Weighs citations by prestige, PageRank-based.
  Pros: Accounts for prestige transfer across citations; Robust against gaming thanks to PageRank algorithm.
  Cons: More complex algorithm; Limited to the JCR database; Favors high-output journals.
- **SJR [7, 18, 29, 38, 49]**
  Description: PageRank-based, weighs citation prestige; Like Eigenfactor considers prestige across citations; Limits but does not eliminate self-citations; Based on Scopus (larger than JCR).
  Pros: Incorporates prestige transfer; Broader database (Scopus).
  Cons: Complex calculation; Limited transparency.
- **SNIP [7, 32, 38]**
  Description: Measures contextual citation impact by weighting citations based on the total number of citations in a subject field.
  Pros: Takes into account the citation potential of the field; Enables fairer comparison across disciplines.
  Cons: Ignores prestige transfer; Less aligned with our goals.

CHAPTER 3

Proposed Approach

Having developed a comprehensive understanding of the relevant metrics, their interrelationships, and the ways in which they can be manipulated, we proceeded to the next phase of our research. In this phase, we aimed to identify the most appropriate metric based on our analysis and objectives. Prior to implementing any automated solutions, we conducted a series of preliminary experiments to explore how data regarding these metrics could be effectively gathered for a given publication. These initial attempts focused on manually collecting the relevant data, allowing us to refine our understanding of the process and assess the feasibility of subsequent automation.

3.1. Early Experiments

In the initial phase of our experiments, we selected a sample of publications and manually retrieved relevant journal metrics from the Scopus, Journal Citation Reports (JCR), and SJR platforms as a proof of concept. Specifically, we gathered a sample of 12 random publications and compiled the data on these publications, including citation counts, into a single CSV file. The data collection process involved identifying the citation counts for each publication and locating the respective International Standard Serial Number (ISSN) of the journal in which the article was published. This ISSN was then entered into the search bar of both the SCImago Journal and Country Rank platform and Clarivate's Journal Citation Reports website. This approach enabled us to retrieve the SJR, SNIP, and H-index from SCImago Journal and Country Rank, even though the SJR and SNIP metrics were also available via Scopus. Additionally, the JIF and Eigenfactor score were extracted from the Journal Citation Reports platform.

This preliminary step provided several key insights. First, it was observed that the Eigenfactor Score was unavailable for certain journals, thereby limiting its utility. Additionally, six journals exhibited values for the SCImago metrics that were accessible via the SCImago platform, yet these same values were not available in the Journal Citation Reports, further corroborating some of the advantages associated with Scopus-based metrics, which draw on a larger database of journals. Inconsistencies were also noted between the H-index and other bibliometric indicators, such as the SJR, Eigenfactor Score, and SNIP. Specifically, in several instances, journals with disproportionately high H-index values displayed relatively low values for the other metrics, particularly in the case of older journals, as illustrated in Table 3.1. These findings resonated with concepts discussed in the literature review, suggesting that these discrepancies warranted further investigation.

TABLE 3.1. Journals with Contrasting H-Index Values

Journal Name	H-Index	SJR	SNIP	ES	JIF
Lecture Notes in Computer Science	470	0.606	0.59	N/A	0.402
IEEE Access	242	0.96	1.44	0.34597	3.4

3.2. First Steps of Automation of Data Gathering

To enable scalable data collection for subsequent analysis, we utilized the system developed in the study "Automatic Detection of Missing Information in the Indexing of Scientific Publications" [**37**]. This system facilitates the identification of missing citation information across various indexing services by allowing users to input a publication's DOI. Building upon this existing infrastructure, our goal was to expand its functionality to include the retrieval of citation data in conjunction with journal-level metrics, rather than merely cataloging the publications and their citing articles. To achieve this, we proposed the development of the following data pipeline:

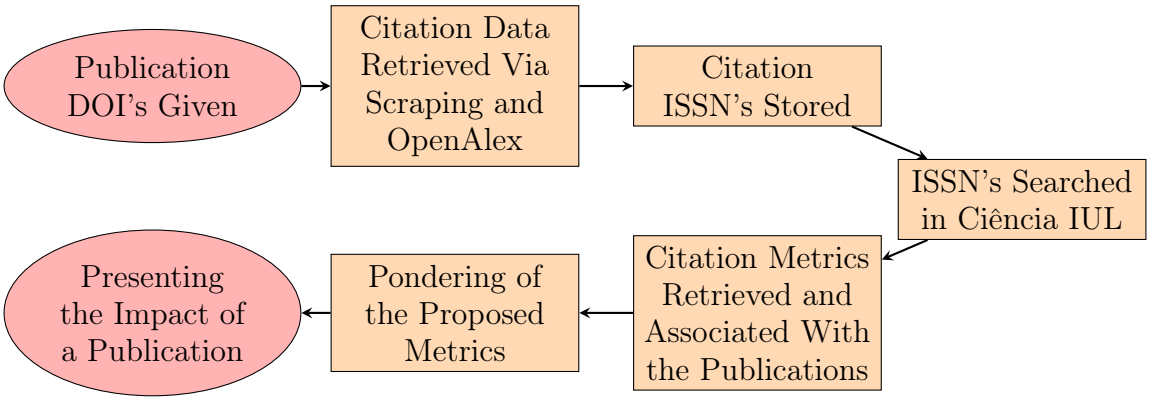


FIGURE 3.1. Data Pipeline Illustration

Given the proposed data collection pipeline, as illustrated in Figure 3.1, we proceeded to establish a robust foundation for automating the data gathering process. Although our approach was built upon the system developed in the work of [**37**], several challenges emerged during its implementation. An initial obstacle was encountered when integrating Scopus into the pipeline. Although the platform did not trigger CAPTCHA validation, unlike in Web of Science, certain interactive elements, such as the export button for downloading citation data in CSV format, contained page-specific unique identifiers. This required additional handling to dynamically interact with these page elements, which was achieved through the use of regular expressions. In contrast, OpenAlex posed no such challenges, providing a comprehensive and openly accessible API that greatly facilitated the retrieval of citation data. Consequently, our system was ultimately designed to extract publication data from both Scopus and OpenAlex exclusively.

Once the data acquisition mechanisms were established, the subsequent step involved retrieving the ISSNs. For Scopus, this was achieved by adding an additional parameter to the Web Scraper designed to select the ISSNs directly from the Scopus website, as

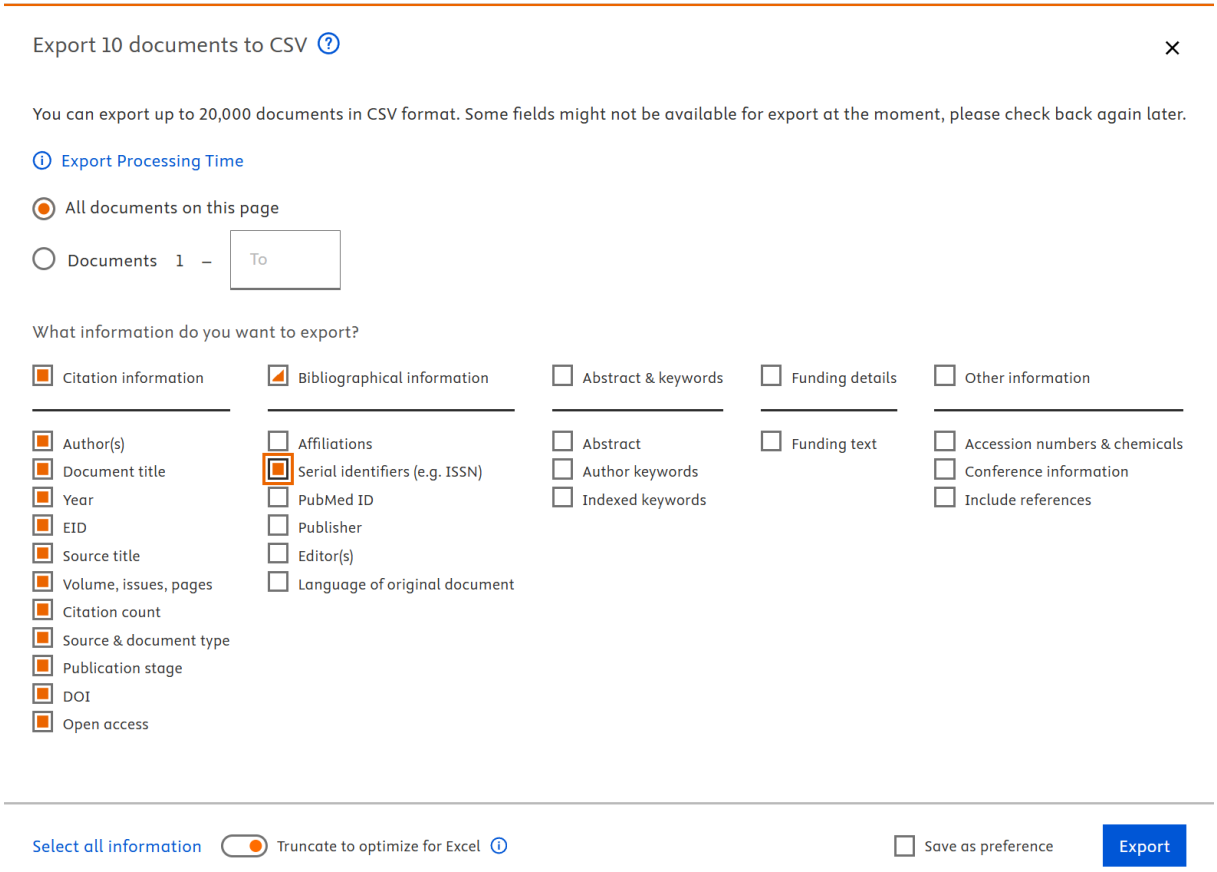


FIGURE 3.2. ISSN Selector

illustrated in Figure 3.2. The next task was to enable the retrieval of the SJR values. This was accomplished by utilizing a CSV file exported from our institution's internal CRIS (Current Research Information System), which provided up-to-date journal metrics and simulated the data retrieval process from Ciência-IUL. All collected data were subsequently stored in a MySQL relational database using Python and the pymysql library. The data were organized into two primary tables: one for Scopus data and another for OpenAlex. These tables contained the following parameters:

- `cited_pub_doi`: The DOI of the cited publications
- `citing_pub_doi`: The DOI of the citation
- `content`: All of the data relating to the citing publication
- `citation_year`: Year when the citing article was published
- `journal_name`: Citing Journal's name
- `journal_issn`: Citing Journal's ISSN and E-ISSN
- `sjr`: The SJR of the citing journal
- `h_index`: The H-Index of the citing journal
- `jif`: The JIF of the citing journal

The three final metric values were initially managed using SQL triggers; however, to facilitate more efficient debugging of the final values, the processing was subsequently handled through Python. As a result, no modifications were made to the database outside

the Python code. A crucial feature of the citation tables was that, among the citing publications, the cited publication also appeared as a citation. This design choice allowed for the inclusion of the impact of the publication's own journal, which is an important step in understanding the broader impact of a publication.

The overarching aim of this project is to investigate whether the true impact of a scientific publication's citations can be meaningfully quantified. For this purpose, the SJR metric was selected as the primary indicator. However, to ensure that the chosen metric accurately reflects the perceived scholarly impact, we decided that the next step would be a user study that involved interviews with domain experts. This approach would ensure that the proposed metrics were validated by experienced researchers who would utilize their own mechanisms for ranking publications based on impact. In this study, participants were presented with a set of publications, along with data regarding the SJR values of the citing journals and additional indicators, such as citation counts and the total SJR sum. They were then asked to rank the publications in order of impact, drawing on their expertise and the provided metrics.

In preparation for the survey, considerable attention was given to optimizing the data visualization. Initially, as depicted in Figure 3.3, the plan was to present the SJR values graphically using a scatter plot, with publication identifiers on the Y-axis and the SJR values of citing journals on the X-axis. However, this approach proved to be suboptimal, as it required participants to infer the number of citations from the graph, which diverted attention from the core metric analysis. As a result, we transitioned to a tabular format for presenting the data, which was first prototyped in Microsoft Excel using exported records from the database, as shown in Figure 3.4. This revised format proved to be significantly more effective, as it allowed for a straightforward inspection of citation data. Moreover, the integrated heatmap enabled immediate visual identification of high and low SJR values, thereby supporting a rapid and intuitive interpretation by survey respondents.

For automation and reproducibility, we developed a system to dynamically generate the citation tables by retrieving publication identifiers (ISSN and E-ISSN) and fetching the associated metrics directly from the previously mentioned database. Initially, Matplotlib was used for data presentation; however, it was later replaced by Plotly due to its more straightforward functionality for generating interactive tables in Python. The interactivity was designed to allow the surveyed researchers to adjust various parameters of the metrics, such as applying a factor to the sum of a metric's values such as $\sum_{i=1} SJR_{pub_i} * factor$. This would enable researchers to fine-tune their preferred metric by making small adjustments, with the goal of ensuring that their ordering preferences were reflected in the metric. However, this interactive feature was ultimately set aside in favor of static metrics, as it offered greater control over what the experts saw during the survey. Subsequently, we asked the participants to indicate which publications did not align with their preferred order.

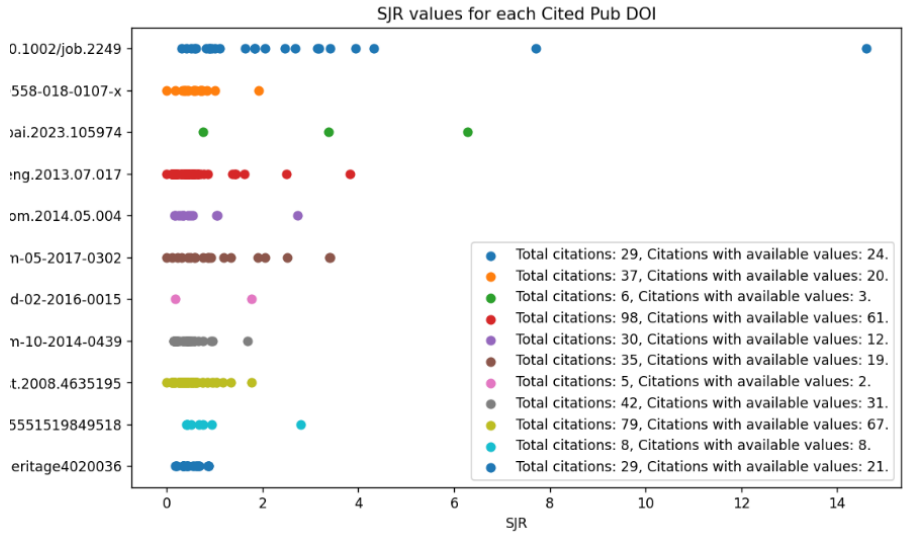


FIGURE 3.3. Visualization of the graph

Rótulos de Linha	Metri	Metri	Metric	Metri															
10.1002/job.2249	60.9	84.9	145.9	24	14.605	7.706	4.321	3.938	3.408	3.187	3.154	2.676	2.473	2.05	1.841	1.841	1.643	1.11	1.001
10.1108/ijchm-05-2017-0302	22.0	41.0	63.0	19	3.408	3.403	2.52	2.05	1.901	1.338	1.196	0.924	0.875	0.86	0.764	0.588	0.581	0.489	0.435
10.1016/j.inpa.2021.04.007	16.3	37.3	53.5	21	2.643	1.091	1.091	1.091	1.008	0.995	0.97	0.915	0.79	0.776	0.61	0.604	0.505	0.485	0.473
10.3390/app10072303	16.2	44.2	60.5	28	2.106	0.96	0.886	0.835	0.825	0.808	0.721	0.703	0.681	0.598	0.508	0.508	0.507	0.507	0.503
10.1108/jkm-10-2014-0439	14.8	46.8	61.6	32	1.693	1.181	0.96	0.927	0.763	0.659	0.583	0.583	0.524	0.508	0.479	0.479	0.438	0.435	0.424
10.1109/taslp.2016.2541299	13.6	34.6	48.2	21	1.959	1.348	1.335	1.138	0.801	0.72	0.72	0.711	0.589	0.524	0.524	0.456	0.435	0.424	0.407
10.1007/s40558-018-0107-x	12.2	32.2	44.3	20	1.916	1.004	0.833	0.74	0.731	0.713	0.708	0.703	0.612	0.575	0.571	0.555	0.435	0.435	0.393
10.1016/j.patrec.2015.12.016	11.2	26.2	37.3	15	4.557	0.745	0.716	0.619	0.525	0.525	0.524	0.522	0.507	0.462	0.411	0.399	0.393	0.252	0
10.1016/j.engappai.2023.105974	10.4	13.4	23.9	3	6.28	3.382	0.763												
10.3390/heritage4020036	10.3	31.3	41.6	21	0.871	0.87	0.855	0.672	0.672	0.628	0.62	0.569	0.544	0.439	0.435	0.415	0.405	0.345	0.344
10.1080/13683500.2022.2115877	10.2	17.2	27.4	7	2.062	1.916	1.763	1.32	1.257	0.967	0.921								
10.3390/app10249082	9.6	25.6	35.1	16	2.137	1.132	1.132	0.764	0.672	0.666	0.617	0.507	0.393	0.339	0.323	0.32	0.238	0.198	0.117
10.1016/j.specom.2017.08.006	9.2	22.2	31.5	13	1.089	1.05	1.05	1.045	1.026	0.971	0.689	0.545	0.524	0.458	0.431	0.181	0.18		
10.3390/info10030094	7.7	19.7	27.5	12	1.348	1.124	1.082	1.069	0.649	0.435	0.435	0.379	0.354	0.353	0.258	0.25			
10.1177/1096348019854793	7.1	16.1	23.2	9	1.618	1.32	1	0.833	0.664	0.543	0.435	0.374	0.314						
10.1177/0165551519849518	7.0	15.0	21.9	8	2.801	0.941	0.761	0.669	0.513	0.434	0.424	0.407							
10.1016/j.specom.2014.05.004	6.8	17.8	24.6	11	2.731	1.05	0.539	0.497	0.435	0.339	0.339	0.295	0.243	0.181	0.157				
10.3390/agriculture12111967	6.7	16.7	23.3	10	1.101	0.95	0.892	0.801	0.688	0.607	0.561	0.435	0.343	0.282					
10.1109/taslp.2018.2820429	6.2	15.2	21.4	9	1.843	1.045	0.689	0.689	0.507	0.435	0.435	0.379	0.165						
10.1108/jd-02-2016-0015	2.1	5.1	7.2	3	1.771	0.176	0.169												
(em branco)																			
Total Geral					**24.384**	**11.727**	**11.421**	**16.764**	**17.12**	**14.725**	**7.455**	**7.022**	**6.476**	**6.548**	**7.62**	**8.431**	**5.419**	**5.032**	**7.596**

FIGURE 3.4. SJR Heatmap Table

With the visualization approach finalized, we compiled a dataset of 20 randomly selected publications from diverse scientific disciplines. The content of the publications themselves was not relevant to the study. Instead, the focus was solely on the impact of the citations. This selection aimed to capture a broad spectrum of impact and prestige, as reflected in the citation profiles, thereby enabling interviewees to examine a variety of metric values and assess the nuanced relationship between citations and scientific impact. A key objective of this sample was to highlight critical discrepancies between publications with high citation volumes and those with perceived high metric values.

All citation data from Scopus were consolidated into a unified DataFrame, which served as the basis for the Plotly table. Citation records from OpenAlex were then merged with the Scopus data, ensuring that only citations with unique DOIs across both sources were retained. Importantly, the SJR values used for analysis corresponded to the year of the citation, not the year of the citing journal's publication, to more accurately

reflect the contemporaneous impact. This adjustment was necessary to correct an issue encountered with some citing articles, which were registered in the Scopus database in one year but associated with a journal edition published in the following year. This mismatch was particularly problematic with the OpenAlex data, which only recorded the year the citations were made available, rather than the publication year of the citing journal.

CHAPTER 4

Development

Building on the insights from the preliminary experiments, we proceeded to develop the final code for the survey implementation. While the initial table presented the values in a satisfactory manner, it was essential for the table to be interactive in order to facilitate sorting. To address this need, we transitioned to using the Dash Python library, specifically utilizing Dash's DataTable component, which enabled value sorting. However, this feature ultimately became unnecessary, as the final version of the table was designed to allow sorting using only HTML and CSS, making the switch from Plotly to Dash somewhat redundant.

As illustrated in Figure 4.1, Dash enabled the creation of an interactive table that effectively displayed a heatmap for the sorted SJR values. Sorting the values was facilitated by a simple JavaScript script incorporated into the HTML code. Another advantage of using Dash was its support for Markdown, which allowed us to embed links to the associated webpages. This feature provided the researchers being surveyed with greater flexibility, allowing them to explore the impact of the publications in the manner they preferred, with all relevant information easily accessible.

With these enhancements, the resulting table offered an intuitive way to inspect the SJR values, thanks to the integrated heatmap. Furthermore, by including links to the publications, researchers could easily comprehend how they would rank the impact of the publications based on the available metrics.

While the primary focus of this study is the SJR metric and its role in understanding the impact of citations on scientific publications, the institution's CRIS system also provides access to the H-index and JIF metrics for journals. Consequently, we sought to

SJR values for each Cited Pub DOI

Cited Pubs	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	Citation 1	Citation 2	Citation 3	Citation 4	Citation 5	Citation 6	Citation 7	Citation 8	Citation 9	Citation 10	Citation 11	Citation 12	Citation 13	Citation 14	Citation 15
Publication 1	29	89.929	333.645	22.870	17.338	14.605	7.706	4.321	3.938	3.408	3.187	3.154	2.676	2.473	2.05	1.841	1.841	1.643	1.11	1.001
Publication 2	36	50.116	106.580	9.218	5.757	3.403	1.916	1.004	0.74	0.731	0.713	0.708	0.703	0.575	0.571	0.555	0.435	0.435	0.393	0.371
Publication 3	6	16.425	58.125	8.849	7.821	6.28	3.382	0.763	0.0	0.0	0.0	_	_	_	_	_	_	_	_	_
Publication 4	25	41.250	106.250	8.683	5.379	2.643	1.091	1.091	1.091	1.008	0.995	0.97	0.915	0.79	0.776	0.61	0.604	0.505	0.485	0.473
Publication 5	20	31.157	75.785	8.164	6.065	4.557	0.745	0.716	0.619	0.525	0.525	0.524	0.522	0.507	0.462	0.411	0.399	0.393	0.252	0.0
Publication 6	29	35.806	63.030	7.234	4.651	2.731	1.05	0.539	0.497	0.435	0.339	0.339	0.295	0.243	0.181	0.157	0.0	0.0	0.0	0.0
Publication 7	15	25.036	65.180	6.735	4.459	1.842	1.089	1.05	1.05	1.026	0.971	0.689	0.545	0.524	0.458	0.431	0.181	0.18	0.0	0.0

FIGURE 4.1. Table With SJR Values

SJR/H-Index/JIF values for each Cited Pub DOI

								Publication			Citation 1			Citation 2			Citation 3			Citation 4			Citation 5		
DOI	Pub Year	#Citations	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF
Publication 1	2017	30	27.687	20.810	631.314	42.394	27.930	3.154	205	4.990	14.605	109	14.900	14.605	109	14.900	7.706	75	16.480	4.321	142	12.230	3.938	205	6.140
Publication 2	2018	39	8.628	4.922	266.195	20.842	13.994	0.740	30	2.190	1.916	108	10.230	1.004	44	2.880	0.833	52	5.200	0.731	30	2.810	0.713	40	3.860
Publication 3	2023	6	8.849	7.821	284.076	29.115	27.256	0.000	0	0.000	6.280	213	25.700	3.382	179	9.220	0.763	49	4.230	0.000	0	0.000	0.000	0	0.000
Publication 4	2021	26	8.767	5.399	560.360	21.840	14.215	1.008	45	7.340	2.643	161	6.650	1.091	193	4.550	1.091	193	4.550	1.091	193	4.550	0.995	33	6.610
Publication 5	2016	20	8.164	6.065	282.307	24.277	20.744	0.000	0	0.000	4.557	136	19.120	0.745	65	4.930	0.716	93	3.110	0.619	88	3.360	0.525	51	3.750
Publication 6	2014	29	7.234	4.651	800.906	12.342	8.801	0.000	0	0.000	2.731	103	4.100	1.050	83	6.100	0.539	78	2.990	0.497	109	1.990	0.435	67	0.840
Publication 7	Unknown	15	6.735	4.459	218.616	15.581	11.185	0.000	0	0.000	1.842	62	7.300	1.089	68	2.240	1.050	83	6.100	1.050	70	3.530	1.026	33	3.080
Publication 8	2022	14	7.110	4.936	205.726	23.727	17.795	2.062	94	10.230	1.916	108	10.230	1.763	58	10.990	1.320	78	6.430	1.257	73	5.900	0.967	45	2.830
Publication 9	2018	35	11.432	7.216	427.914	33.658	23.128	0.000	0	0.000	3.408	172	13.380	3.403	159	8.860	2.520	67	15.120	2.050	70	9.120	1.901	39	8.000
Publication 10	Unknown	7	3.952	3.132	75.624	8.117	7.233	0.000	0	0.000	1.771	19	5.940	0.479	31	2.260	0.176	13	0.760	0.169	63	0.000	0.000	0	0.000

FIGURE 4.2. Table with three columns per citation

present these metrics alongside the SJR to provide a more comprehensive view of the available bibliometric data, ensuring that these metrics were not hidden behind publication links.

To facilitate this, modifications were made to the table. The citation columns were expanded to include each of the metrics, and the publication links were separated from the metrics themselves, resulting in the revised table shown in Figure 4.2. Although this approach slightly reduced the immediate interpretability of the data, it allowed researchers to directly view some of the most important metrics without needing to open any publication or citation links. This design choice aimed to provide a more transparent presentation of the metrics, thus minimizing the potential for bias that could arise from presenting only the SJR values. It could be argued that presenting a single metric alone might influence researchers to rank publications based on that metric, given that it would be the only immediately visible data.

Only the first 30 citation's metrics were presented, as beyond the 30th citation, the metric values are typically very low, rendering the omitted citations largely insignificant for assessing the prestige of the publications. Limiting the displayed citations to this subset also ensures that, in cases where a publication has a very high citation count, the table remains at a manageable and readable size. It is important to note, however, that although only 30 citations were shown, all citations were considered when calculating the proposed metrics.

In addition, the publication itself was separated from the citation data, with publication years clearly displayed. This further contributed to a more comprehensive understanding of the metrics, ensuring that interviewees had all the relevant information readily available without needing to navigate external resources.

								Publication			Citation 1			Citation 2			Citation 3			Citation 4			Citation 5		
DOI	Pub Year	#Citations	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF
Publication 1	2017	29	30.372	21.975	763.231	48.889	31.397	3.154	205	4.990	14.605	109	14.900	14.605	109	14.900	10.486	306	17.090	7.706	75	16.480	4.321	142	12.230
Publication 2	2018	37	11.968	7.456	509.679	30.894	20.278	0.740	30	2.190	3.403	159	8.860	3.383	236	12.480	3.068	236	8.680	2.228	201	4.920	2.217	151	7.020
Publication 3	2023	5	9.562	8.031	344.652	32.474	28.181	1.749	137	9.220	6.280	213	35.700	3.382	179	9.220	1.749	137	9.220	0.763	49	4.230	0.000	0	0.000
Publication 4	2021	26	8.886	5.434	563.670	22.089	14.275	1.008	45	7.340	2.643	161	6.650	1.091	193	4.550	1.091	193	4.550	1.091	193	4.550	0.995	33	6.610
Publication 5	2016	19	9.311	6.567	524.278	28.963	22.458	0.729	170	3.160	4.557	136	19.120	2.070	249	9.430	1.875	271	9.290	0.848	170	4.270	0.745	65	4.930
Publication 6	2014	46	8.762	5.097	820.216	15.272	10.132	0.605	113	2.330	2.731	103	4.100	1.050	83	6.100	1.045	70	5.280	0.546	113	2.870	0.539	78	2.990
Publication 7	2017	17	7.515	4.779	301.538	18.409	12.390	0.546	113	2.870	1.842	62	7.300	1.089	68	2.240	1.050	83	6.100	1.050	70	3.530	1.045	70	5.280
Publication 8	2022	13	11.012	7.428	401.592	38.540	26.363	2.062	94	10.230	2.923	169	13.240	2.923	169	13.240	2.923	169	13.240	2.843	126	11.630	2.843	126	11.630
Publication 9	2018	35	15.215	8.929	629.227	45.736	28.126	1.849	113	4.510	3.408	172	13.380	2.843	126	11.630	2.594	191	10.270	2.520	67	15.120	2.512	151	9.380
Publication 10	2016	6	4.546	3.480	138.557	9.372	7.797	0.696	69	1.270	1.771	19	5.940	0.836	69	2.970	0.836	69	2.970	0.169	63	0.000	0.000	0	0.000

FIGURE 4.3. Finalized Table with ISSN Values Corrected

This approach allowed us to gather insights into how researchers compare their expert opinions on the publications. These insights are critical for understanding which of the available metrics most effectively reflect the experts' rankings.

At the time of data collection for the tests, the code operated as follows: the group of publication DOIs was submitted through the code, which then checked the database to determine if the DOIs were present and whether the associated values were older than thirty days, as outlined in [**37**]. If the values were absent or more than thirty days old, new values would need to be retrieved. With the modifications to the data-fetching process through web scraping, the Scopus values were freely available. However, at the time of writing (16/09/2025), a user who is not signed into Scopus is now unable to export any data from a list of publications, such as the citation list for a given publication. As a result, an alternative workaround will need to be developed to circumvent this limitation.

However, during this process, the table revealed several issues related to the ISSN and E-ISSN values. In many instances, the ISSNs of the publication journals and citations were incorrectly recorded. This was often due to mismatched ISSNs, especially when the ISSNs that began with a '0' were omitted, leading to errors in the values stored from Scopus and OpenAlex. There were also cases in which the ISSN and E-ISSN values were swapped. To address these issues, the method for retrieving these values was revised.

Upon review, we observed that values for publication 3's journal were now correctly available, rather than being marked as 0. This issue stemmed from the fact that the ISSN of that publication began with a '0', which was removed in the CRIS database CSV, causing the value not to be found. However, it was found that not all ISSN that start with 0 would have that removed.

Initially, our first approach was to remove the leading zero from any ISSN values being searched and to surround the ISSN in the SQL query with wildcard characters before and after. This ensured that, if too many zeros were removed, the use of wildcards would still allow the correct value to be matched. However, this created issues with certain

journals. For example, the journals Speech Communication and Musica Hodie have ISSNs 01676393 and 16763939, respectively. Since the ISSN of Speech Communication began with a zero, this value was missing in the database, resulting in the stored ISSN being 1676393. When combined with wildcards on both sides, this value incorrectly matched Musica Hodie's ISSN, causing the query to return results for Musica Hodie instead of Speech Communication.

To resolve this, we restricted the query to use only one wildcard character, placed before the ISSN. This adjustment allowed for ISSNs with missing leading zeros to still be found, while ensuring that any ISSN with differing values would be correctly identified as distinct. The final query was therefore defined as follows:

```
SELECT SJR, h_index, cites_docs_2years_ratio, ref_year
FROM ciencia_data
WHERE issn LIKE %journal_identifier
```

This query was run using combinations of ISSN and E-ISSN, with wildcards placed before the identifier to account for possible discrepancies in the registration of ISSNs and E-ISSNs. The approach was designed to accommodate minor errors in the ISSN/E-ISSN entries without modifying the registered journal data. The main limitation of this query is the potential for false positives. For example if Musica Hora's ISSN was 16763939 and not 1676393, there would still be a false positive when searching for Speech Communication. We experimented with searching for the ISSN and the name of the journal, but there were too many discrepancies to make this viable. Additionally, it is important to note that hyphens separating the ISSN components were removed, concatenating the two parts of the ISSN and E-ISSN.

4.1. The Metrics

All the data presented in the table serves to offer as much information as possible to the researchers, enabling them to evaluate the publications based on their individual evaluation criteria. The following alternative metrics were proposed to provide a broader representation of the three indicators:

- Metric 1: SJR-Adjusted Citation Impact (Quadratic)

$$\sqrt[2]{\lambda(SJR_p + 1)^2 + \sum(SJR_i + 1)^2}$$

- Metric 2: SJR-Adjusted Citation Impact (Cubic)

$$\sqrt[3]{\lambda(SJR_p + 1)^3 + \sum(SJR_i + 1)^3}$$

- Metric 3: H-index-Adjusted Citation Impact (Quadratic)

$$\sqrt[2]{\lambda(HIndex_p + 1)^2 + \sum(HIndex_i + 1)^2}$$

- Metric 4: JIF-Adjusted Citation Impact (Quadratic)

$$\sqrt[2]{\lambda(JIF_p + 1)^2 + \sum(JIF_i + 1)^2}$$

- Metric 5: JIF-Adjusted Citation Impact (Cubic)

$$\sqrt[3]{\lambda(JIF_p + 1)^3 + \sum(JIF_i + 1)^3}$$

For simplicity, in the remainder of this document these metrics will be referred to by their shorter labels: *Quadratic SJR*, *Cubic SJR*, *Quadratic H-Index*, *Quadratic JIF*, and *Cubic JIF*.

In addition to providing the citation metrics for researchers, we aimed to present alternative metrics that offer a more comprehensive representation of the three indicators. These metrics are inspired by the Euclidean distance formula, with the indicator values augmented by a base flat value, which serves as a minimum for the citation impact. This ensures that the prestige of a citation is reflected in the overall citation impact.

By introducing a flat base for all citations, even those with low indicator values retain some measure of impact. At the same time, citations from highly prestigious journals are emphasized due to the nonlinear nature of the equations. For the SJR and JIF indicators, we applied a cubic version of the metrics to give greater weight to higher indicator values. This adjustment was deemed unnecessary for the H-Index, as its value range is already significantly larger compared to the other metrics.

The metric that we propose as the ideal is the first one in the list, which we have termed the SJR-Adjusted Citation Impact Metric. As supported in the literature review, the SJR is the most balanced, fair, and robust indicator among the three widely available metrics. The first metric provides a more nuanced view of the publications and the impact of their citations. It allows lower citation publications to benefit from the prestige of their citations, while not diminishing the value of highly cited publications with lower impact citations.

Thus, the resulting ranking of publications based on this metric balances both highly cited and lower-cited publications in a way that reflects the complexities and nuances of the publications' true impact.

The final ranking of each proposed metric and the number of citations are as follows:

TABLE 4.1. Metric Rankings Across Publications

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
#CIT	11	6	2	9	19	17	1	4	12	18	5	7	15	20	16	8	13	14	10	3
M1	1	9	2	8	11	3	5	17	4	6	15	12	19	7	18	20	16	13	14	10
M2	1	9	3	2	8	5	15	11	4	17	12	6	7	18	19	14	13	20	16	10
M3	6	1	11	9	12	18	4	5	2	14	17	8	3	13	15	19	20	7	16	10
M4	1	9	8	3	2	5	15	17	18	12	11	4	7	19	13	14	16	6	20	10
M5	1	3	9	8	5	2	15	18	17	12	14	11	4	13	7	19	16	6	20	10

4.2. Visualization of the Data

One initial answer to our researcher survey stood out because of its thorough description of how that researcher ordered the metrics themselves. What follows are their orderings:

The metric identified as most representative was the Quadratic JIF, owing to its close alignment with the researcher's ordering. Nonetheless, the researcher observed that both

TABLE 4.2. Respondent's Publication Ranking

Rank	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th
Pub	17	20	13	7	4	9	8	6	5	3	12 or 1	18	14	16	2 or 11	19	10

the Quadratic SJR and the Cubic JIF also demonstrated considerable similarity, with six publications occupying identical ranking positions to the researcher's list, compared with seven in the case of the Quadratic JIF. This resemblance is not coincidental, as established in Section 2.4, both SJR and JIF exhibit similar behavior. The researcher's detailed response, together with prior knowledge of the similarities and distinctions among these metrics, motivated us to undertake a visualization and comparison of their behavior, including that of the proposed metrics. Accordingly, before proceeding with the analysis of survey responses, several visualization exercises were conducted, leading to preliminary conclusions regarding the behavior of these metrics.

Initially, we aimed to identify patterns and notable observations simply by sorting the table's values. A simple example of this is shown in Figure 4.4, where, due to the heatmap representation, it becomes evident that the SJR and JIF values, even across different ranks of citations, do not differ drastically. It is important to note that comparisons across the citations do not provide a true ordering of the impact of the publications in the rows. This has been a point of confusion given the nature of the table. The columns of the table represent the nth rank of the highest SJR citation corresponding to the respective publication in its row, as the table rows are sorted by the SJR values. This sorting method can be misleading, as there is no inherent connection among the citations within a column. They appear in that column solely due to their SJR rank.

However, despite the limitations of this sorting method, some preliminary patterns can still be identified. The rows are ordered based on the descending order of the SJR values, specifically the second-highest SJR value of the citations. While this sorting method does not provide the clearest depiction of the impact of the publications, some patterns emerge across the columns. For instance, a straightforward comparison can be made between the SJR and JIF values, which generally follow similar trends like those observed in Section 2.4. The only exception to this is observed in the first citation column, where the values are more disparate, breaking the generally smooth gradient of the heatmaps across the different columns.

In the second citation column, the JIF values exhibit a downward trend that closely mirrors the downward progression of the SJR values. Even across the citation columns, the descent of values generally follows a similar pattern for the two metrics. This trend is in stark contrast to the H-Index values, which appear more erratic, with considerable fluctuations. Nonetheless, there is an observable, albeit inconsistent, downward trend in the H-Index values. The key observation here is the significant discrepancy between the behavior of the SJR and JIF values and that of the H-Index values.

This observation, again, aligns well with the observations made in Section 2.4. Since the rows are ordered based on a single citation column from the different publications, no

meaningful correlation can be established between the ordering of rows and the metrics 1-5. This is because the metrics tally all the indicator values for each publication, which means that a publication with numerous lower-impact citations may accumulate a higher score in the metric system, resulting in a higher ranking. This, in turn, contrasts with the SJR-based sorting shown in the image, where the publication's rank is influenced primarily by the highest citation values.

								Publication			Citation 1			Citation 2			Citation 3			Citation 4			Citation 5			Citation 6		
DOI	Pub Year	#Citations	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	SJR	H	JIF	SJR	H	JIF	SJR ▼	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF	SJR	H	JIF
Publication 1	2017	30	30.372	21.975	763.231	48.889	31.397	3.154	205	4.990	14.605	109	14.900	14.605	109	14.900	10.486	306	17.090	7.706	75	16.480	4.321	142	12.230	3.938	205	6.140
Publication 2	2018	38	11.968	7.456	509.679	30.894	20.278	0.740	30	2.190	3.403	159	8.860	3.383	236	12.480	3.068	236	8.680	2.228	201	4.920	2.217	151	7.020	1.916	108	10.230
Publication 3	2023	6	9.562	8.031	344.652	32.474	28.181	1.749	137	9.220	6.280	213	15.700	3.382	179	9.220	1.749	137	9.220	0.763	49	4.230	0.000	0	0.000	_	_	_
Publication 8	2022	14	11.012	7.428	401.592	38.540	26.363	2.062	94	10.230	2.923	169	13.240	2.923	169	13.240	2.923	169	13.240	2.843	126	11.630	2.843	126	11.630	1.916	108	10.230
Publication 9	2018	36	15.215	8.929	629.227	45.736	28.126	1.849	113	4.510	3.408	172	13.380	2.843	126	11.630	2.594	191	10.270	2.520	67	15.120	2.512	151	9.380	2.512	151	9.380
Publication 15	2019	16	8.479	5.752	323.741	25.216	18.283	1.618	84	4.640	2.928	151	13.620	2.288	113	7.680	1.568	120	7.480	1.402	88	7.440	1.320	78	6.430	1.000	56	3.830
Publication 5	2016	20	9.311	6.567	524.278	28.963	22.458	0.729	170	3.160	4.557	136	19.120	2.070	249	9.430	1.875	271	9.290	0.848	170	4.270	0.745	65	4.930	0.716	93	3.110
Publication 12	2016	25	8.314	5.113	602.860	23.480	15.658	0.711	70	4.080	1.959	171	8.780	1.723	210	7.320	1.348	70	5.980	1.335	70	6.170	0.801	106	4.530	0.720	93	4.270
Publication 11	2015	49	10.182	5.503	666.761	22.826	14.346	1.181	134	3.520	1.693	101	9.100	1.397	92	7.590	0.960	242	4.640	0.927	204	4.440	0.763	81	3.960	0.763	81	3.960
Publication 18	2020	21	7.375	4.699	586.198	24.122	18.209	0.435	101	2.990	2.137	34	15.530	1.132	86	4.900	1.132	86	4.900	1.132	86	4.900	0.764	219	4.420	0.672	169	3.950

FIGURE 4.4. Image of the table being sorted by SJR

This demonstrates that noticeable trends and patterns can indeed be identified in the metrics presented in the table. These patterns and trends appear to align with some of the concepts discussed in Section 2.4. In an attempt to gain further insight into these patterns, we created several visualizations to determine if these trends could be replicated. To begin, we constructed three scatter plots comparing the SJR and JIF, SJR and H-Index, and H-Index and JIF.

As illustrated in Figure 4.5, there is a strong relationship between the SJR and JIF values, with a few exceptions, such as the first two citations from publication 1, where the SJR values are significantly higher than the corresponding JIF values. The comparison between SJR and H-Index is slightly harder to interpret, as the points appear more dispersed. Nevertheless, while the visualization shows clustering at the lower range of H-Index values, the regression line suggests a slight positive relationship, although it is much weaker than the one observed between SJR and JIF. Lastly, the comparison between JIF and H-Index demonstrates a clearer positive relationship compared to SJR vs H-Index. This suggests that journals with higher H-Index values also tend to have higher JIFs. However, the dispersion of points, particularly at the higher H-Index ranges, indicates that these metrics are not perfectly aligned and may be capturing different aspects of scholarly impact.

In summary, the scatter plots suggest that SJR and JIF are strongly related and can be used somewhat interchangeably in this context, while the H-Index behaves differently,

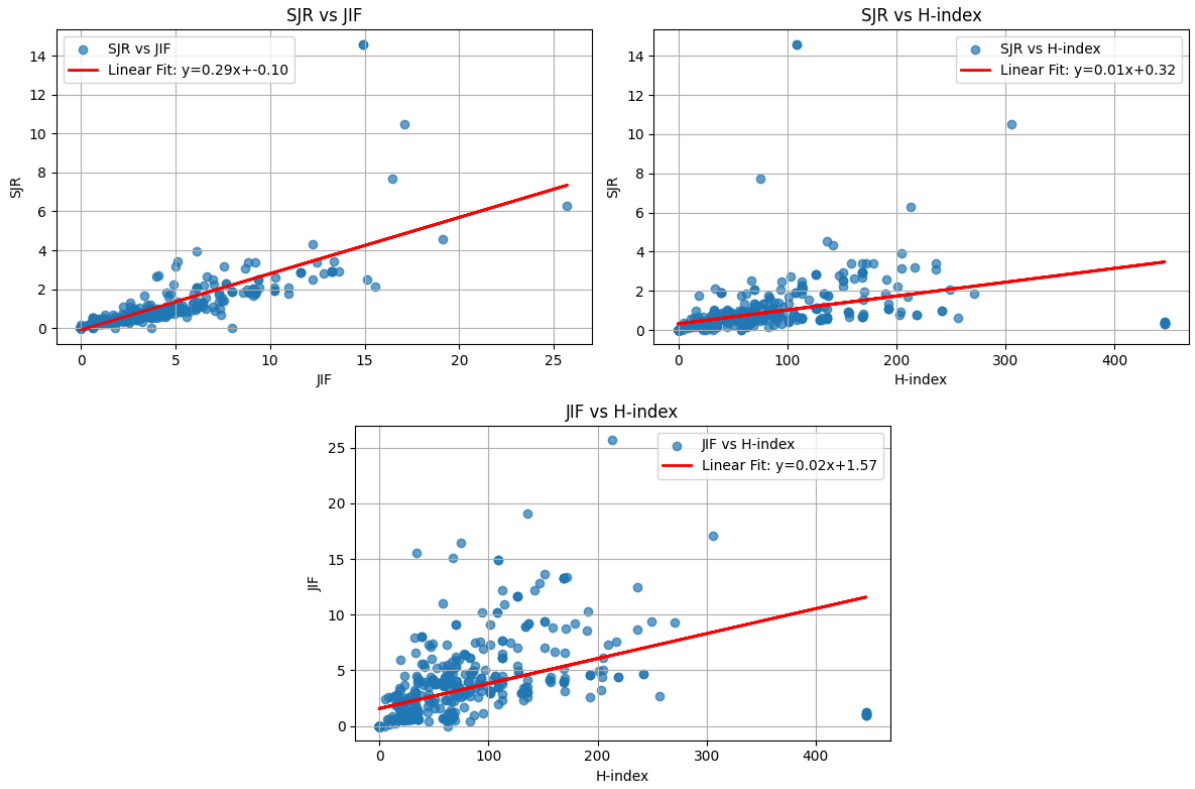


FIGURE 4.5. Relationship Between Metrics

correlating weakly with SJR and moderately with JIF. This reinforces the notion that the three indicators, while related, capture distinct dimensions of journal influence.

Next, we created a rank plot graph that shows the values of SJR, H-Index, and JIF to assess how these values align. Given that there are 20 publications, first, we limited the graph to only the first 5 publications to improve interpretability. It is important to note that the metrics will be displayed ordered by the citation rank of the SJR values in descending order. Thus, the JIF and H-Index values represent the metric values for the citations, ordered according to the descending SJR values.

At a surface level, it is apparent from the plot in Figure 4.6 that all the metrics exhibit a general downward trend. However, the H-Index shows significantly more fluctuations, particularly with higher-ranked publications that have lower citation impact. This variability is further exacerbated by the scale of the H-Index, which ranges from 0 to 446, making the graph disproportionately scaled when compared to the values of SJR and JIF. To address this issue, we created another plot featuring only the SJR and JIF metrics to better understand their trends across the ranks, since we know from Section 2.4 these two should correlate well.

The comparison in Figure 4.7 provides a clearer perspective, as the values of SJR and JIF are more closely aligned. While the JIF does exhibit some fluctuations that are inconsistent with the progression of SJR values, these discrepancies may be attributed to the JIF's susceptibility to citation manipulation, a concern not as pronounced in the

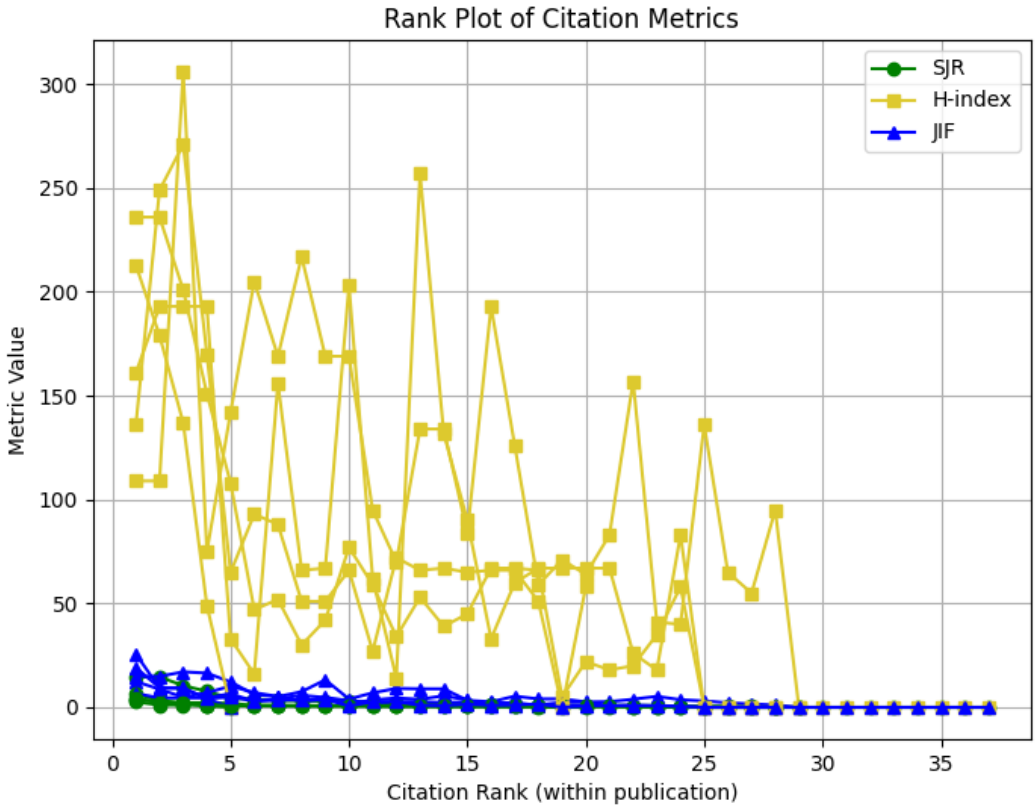


FIGURE 4.6. Rank Plot of SJR H-Index and JIF

case of SJR. Despite these fluctuations, the majority of publications generally align closely with the SJR values across the various ranking tiers. This prompted us to fit both linear and logarithmic regression lines to the data, as this would provide a clearer understanding of how these metrics behave. The choice of a linear regression was made to offer an intuitive trend line, reflecting a general idea of how the metrics evolve across the citations. Meanwhile, the logarithmic regression was chosen to reflect the diminishing trend of citations along the x-axis and to examine how the metrics behave within the context of the sorted SJR values.

The regression lines in Figure 4.8 suggest that the distributions of SJR and JIF evolve similarly, despite the different scales of the metrics. This led us to further investigate these trends by including all available publications, rather than just the top five. This was particularly important due to the presence of zero values in the metrics. For instance, one of the publications in Figure 4.8 had 12 citations but with zero values for both SJR and JIF, which could skew the regression results by not providing sufficient data to represent the trends accurately.

In Figure 4.9, the regression line coefficients for SJR and JIF do not exhibit substantial differences. Although these two metrics differ considerably in their orders of magnitude,

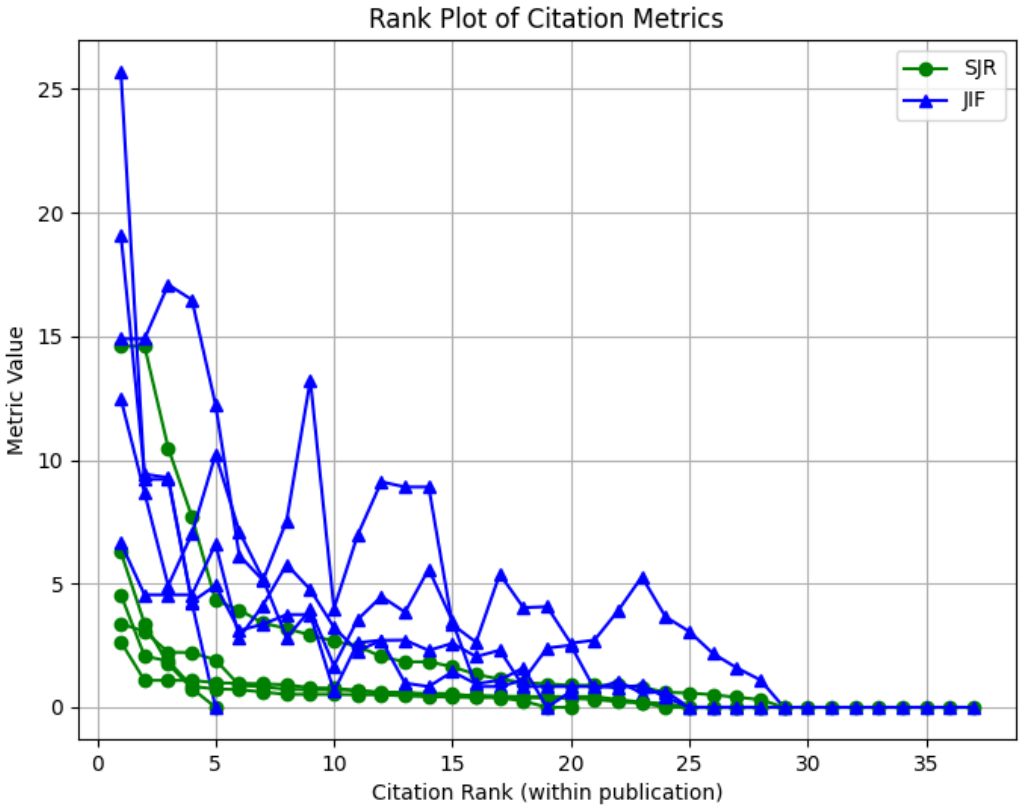


FIGURE 4.7. Rank Plot of SJR and JIF

both demonstrate similar downward trends and comparable behavior within the logarithmic regression analysis. Nevertheless, even when incorporating all citations into the graph, a noticeable discrepancy persists. This disparity arises primarily from the lack of normalization, as the differences in scale between the metrics impede direct comparison. To address this limitation, the metric values were normalized to a scale ranging from 0 to 1, following the approach illustrated in Figure 1 of [**18**]. Such normalization enables a more accurate and meaningful comparison between the metrics.

For Figure 4.10, we reintroduced the H-index. Since the values are now normalized, the order of magnitude between the values of the metrics becomes irrelevant, thereby simplifying visualization and providing the most accurate representation of how closely the values are distributed. As illustrated by the linear and logarithmic regressions, the SJR and JIF metrics exhibit very similar distributions, with a minimal difference in the line coefficients: 0.001 for the linear regression and 0.003 for the logarithmic regression. This indicates that the distribution of these metrics is strikingly similar over the course of citations across the publications. This further supports the notion that, even with the relatively small 20-publication dataset gathered across multiple fields, these two metrics, which differ both in their calculation and in their interpretation of scholarly impact, display comparable behaviors.

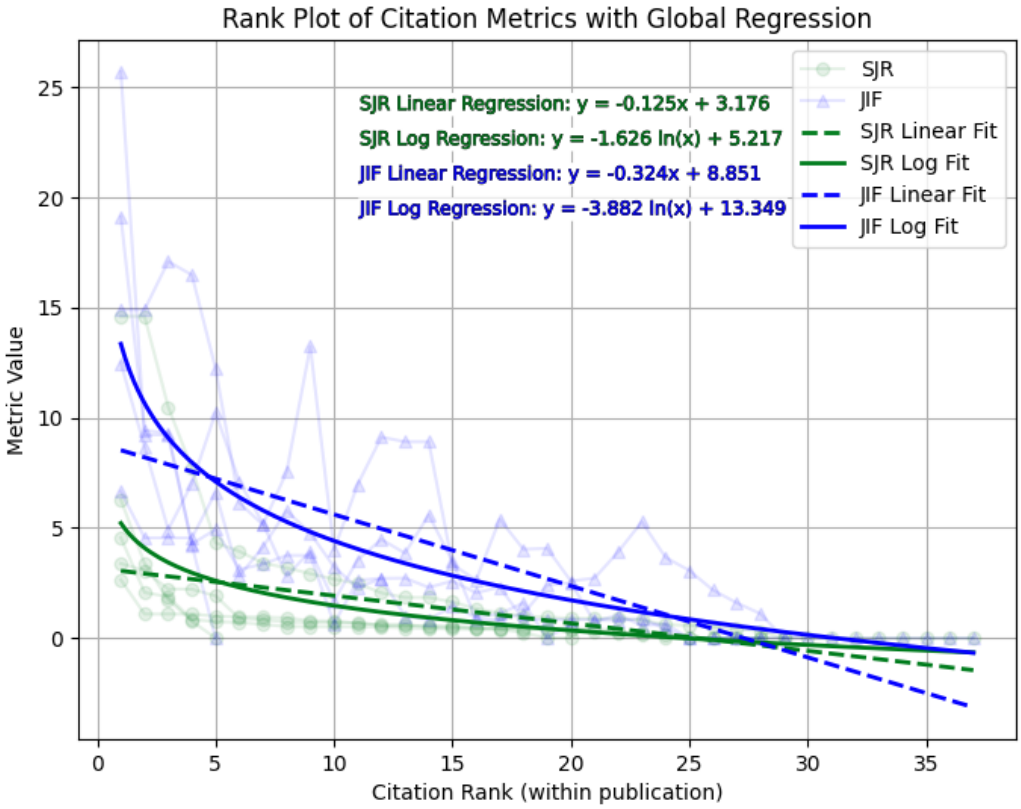


FIGURE 4.8. SJR and JIF Behaviours

In contrast, the logarithmic regression for the H-index is notably different from that of SJR and JIF, although the linear regression is quite similar, with a mere 0.003 difference in the line coefficient compared to SJR. This suggests that, while all metrics show similar trends over the course of citations, the differences in the logarithmic regression highlight disparities in how these metrics behave along the citation ranks. The very small difference between the SJR and JIF line coefficients suggests that these metrics maintain a similar sensitivity to variations in values, implying that the JIF values, as a whole, do not deviate significantly from SJR values, an observation that may not have been immediately apparent, given their differing methodologies. The JIF is based on average citations within a two-year window, whereas SJR uses a network-based approach to measure the transfer of impact over a three-year period, incorporating various mechanisms to limit citation malpractice. Despite these differences in calculation and interpretation, the two metrics exhibit similar behaviors, almost appearing interchangeable if used to order the publications, akin to what is being done for the graphs. This outcome aligns with the findings in Section 2.4. As previously hypothesized, while the H-index trends in a manner similar to these citation-impact metrics, it diverges in behavior, prioritizing a different interpretation of impact that more heavily favors long-term productivity, balanced with

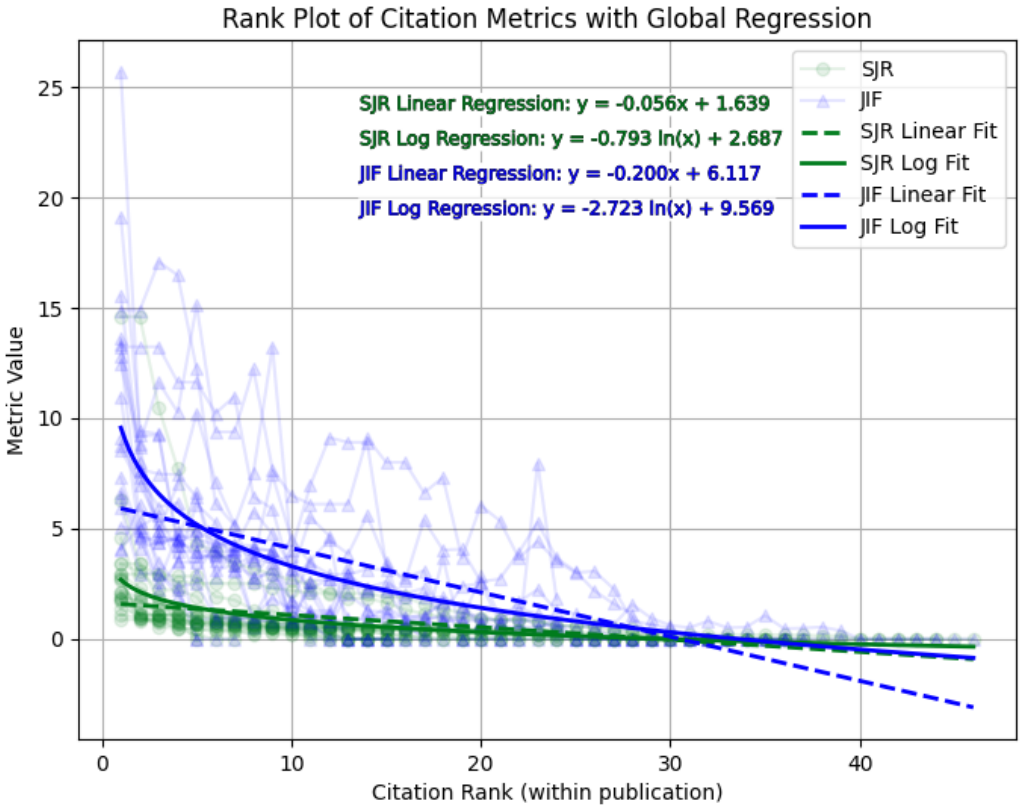


FIGURE 4.9. Behaviors With All Available Data

popularity, as opposed to focusing on the impact of each individual publication across different journals.

While these conclusions provided some insight, the distinction and interpretation of these patterns and behaviors could be better represented. A more effective approach to presenting these relationships would involve correlation coefficients, particularly those explored in earlier sections, such as Pearson and Spearman correlations.

Although the data in Figure 4.11 may not be immediately intuitive, the heatmap provides a valuable understanding of how these metrics correlate, aligning well with the correlation coefficients presented in Section 2.4. As expected, the SJR and JIF metrics demonstrate a strong correlation with one another and a weaker correlation when compared to the H-index. Notably, the Spearman correlation values tend to be higher than the Pearson correlation coefficients, which correlates with the findings from the previous normalized rank plot in Figure 4.10. The high similarity in the linear regression coefficients between the different metric pairs is consistent with the higher Spearman correlation values, as Spearman's correlation measures monotonic relationships, which is appropriate given the downward trend of all metrics. This justifies the high correlation values across the board, with SJR and JIF consistently showing the strongest correlations. In contrast, the Pearson correlations align with the coefficients of the logarithmic regressions,

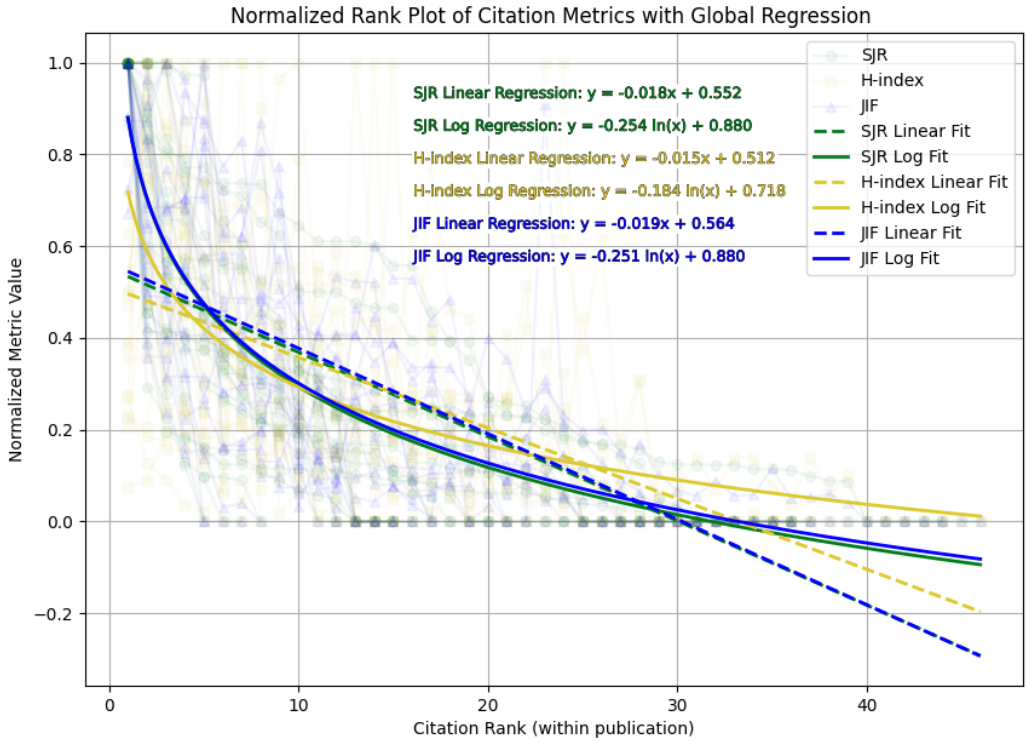


FIGURE 4.10. Metric Behaviours Normalized

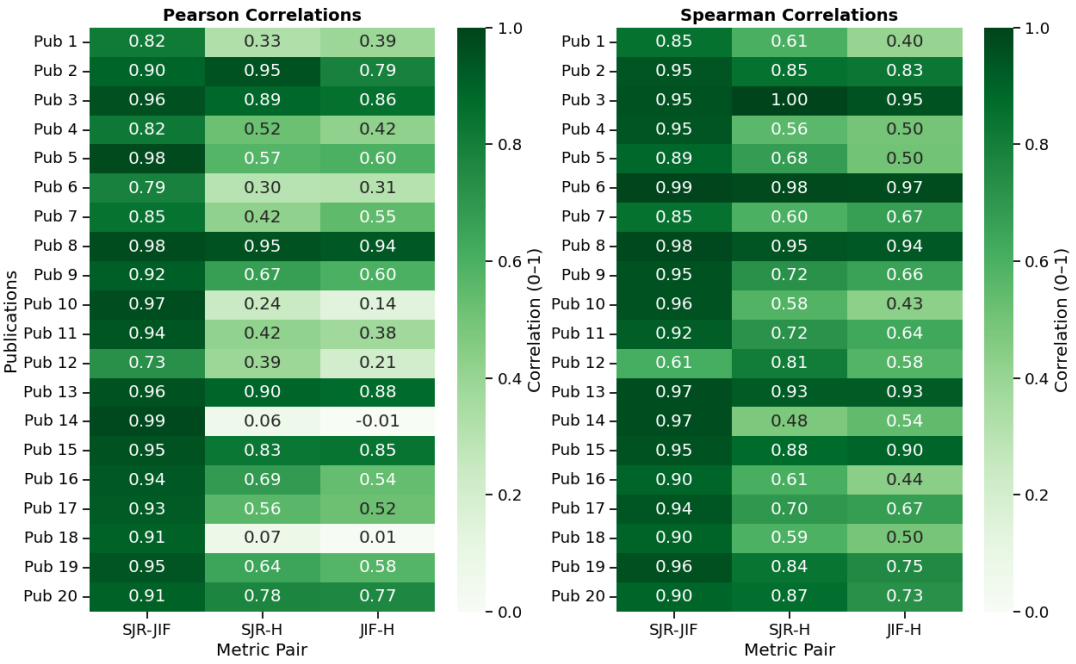


FIGURE 4.11. Citation Data Correlations

reflecting the more subtle differences in values along the citation ranks. Again, the SJR and JIF metrics are highly consistent in their behavior, whereas the SJR-H-Index and JIF-H-Index correlations are notably lower, even approaching zero in some cases.

At a more granular level, a few publications stand out. Notably, publication 3 has a perfect Spearman correlation score of 1 between SJR and H-Index and generally shows higher-than-average correlation values. This is likely due to the fact that publication 3 only has six citations, meaning that the limited data points yield a more consistent trend, which invalidates any interpretation that can be made with values as the population of data is far too small to make any conclusion. Another noteworthy publication is pub 14, which exhibits an almost perfect Pearson correlation coefficient but extremely low values for both SJR and H-Index, as well as for JIF and H-Index. This is also reflected in the Spearman correlations, though with slightly higher values. Additionally, publications 8, 13, and 15 display the highest correlation values among all publications. However, upon inspecting the table, these publications do not exhibit any anomalous behavior and appear to simply have strong correlations.

This more granular examination of the values offers the most intuitive and comprehensive way of understanding the metrics within the context of our study. However, one final presentation of these values will complete the interpretation and further strengthen the connection between our findings and those in Section 2.4. This final step involves calculating the average of the correlation coefficients for each publication.

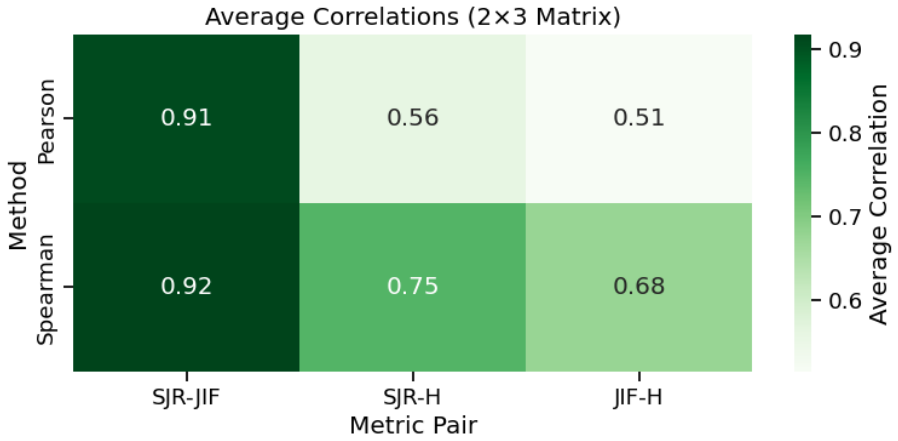


FIGURE 4.12. Average Correlations

This approach allows us to compactly observe the very high correlations between SJR and JIF, and how these correlations significantly decrease when the H-Index is compared with the two other metrics, as shown in Figure 4.12. Once again, the correlation values indicate a weaker relationship between the SJR-H-Index and JIF-H-Index pairs, which aligns with the small difference in line coefficients observed in the linear regression analysis and the significantly lower Pearson correlation coefficients. This observation further corresponds to the difference in slope observed in the logarithmic regression.

In summary, the hypothesized correlation between the SJR and JIF metrics, as discussed in our literature review, is largely confirmed by the values in our comparative table. This suggests that the ordering of publications based on SJR and JIF is quite similar.

Since we intend to ask researchers to rank the publications in the table according to their own impact assessment mechanisms, these correlations could provide valuable insights. Comparing the researchers' rankings with the metric they find most representative of their ordering will allow us to explore how the similarities in behavior between SJR and JIF metrics align with their own rankings.

Before proceeding to this analysis, however, it is important to assess how well our proposed metrics correlate with the existing ones. Since these metrics were designed to magnify the value of a citation based on its journal's impact, it is possible that there could be some unanticipated skewing of the values, potentially leading to unexpected outcomes. This step is crucial in determining whether the behavior of these metrics aligns with the intended representation of citation impact.

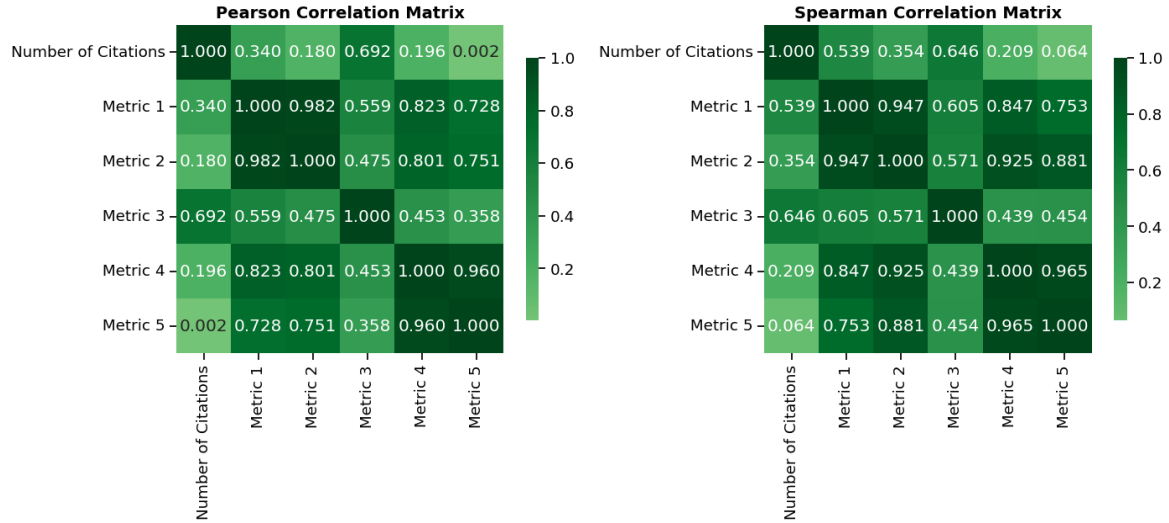


FIGURE 4.13. Pearson and Spearman Correlations of Proposed Metrics

Figure 4.13 illustrates a heatmap displaying the Pearson and Spearman correlation coefficients between the selected metrics and the number of citations. The results clearly reveal high correlation coefficients between Metrics 1 and 2, as well as between Metrics 4 and 5. This outcome is to be expected, as these metrics are derived from the same underlying indicators: SJR for Metrics 1 and 2, and JIF for Metrics 4 and 5.

Another noteworthy observation is the consistently high correlation between the proposed metrics 1 and 4, 1 and 5, 2 and 4 and 2 and 5. This finding is consistent with the previously established strong correlation between SJR and JIF, as discussed in [**18**]. Given the cross-disciplinary nature of the publications selected for this work, it is insightful to compare these correlation values with global values reported in Table 2 of the aforementioned work.

For the Pearson correlation between Metrics 1 and 4, the quadratic transformations of the formulas yield the highest values. These proposed metrics were specifically designed to accentuate the weight of a citation in conjunction with the available metric values.

Consequently, a high correlation between these calculations (Metrics 1 and 4) is expected, since we know the SJR and JIF correlate well with each other.

An important observation arises when considering why comparisons involving the cubic transformations of these metrics result in slightly lower correlations. This can likely be attributed to the fact that the quadratic formulation achieves a balance between the value of a citation and that of the metrics, whereas the cubic formulation disproportionately amplifies the role of the metrics. In doing so, the cubic version diminishes the relative contribution of the citation itself. The quadratic formula, on the other hand, levels the contributions of both factors, ensuring that the citation itself retains significant weight in the calculation, rather than allowing the metrics alone to dominate, hence a closer linear alignment.

The highest correlation value across both matrices is observed in the very strong Spearman correlation coefficient between Metrics 2 and 4, corresponding to the cubic transformation of the SJR metric (designed to strongly accentuate SJR values) and the quadratic transformation of the JIF metric. This result indicates that discrepancies between the two rankings are minimal and that both proposed metrics follow highly similar trends. Although their Pearson correlation is slightly weaker (0.801), the Spearman coefficient reaches 0.925. This divergence can likely be attributed to the cubic transformation of the SJR metric. Since the cubic function disproportionately emphasizes large metric values, and because SJR and JIF are measured on different numerical scales, the strict linear relationship weakens while the relative ranking remains highly consistent. Furthermore, as the raw values of these metrics are already known to exhibit strong correlation, the cubic transformation of SJR, whose untransformed values are measured on a lower scale than those of JIF, creates a scenario where the SJR values follow the JIF values more closely. In other words, the numerically lower SJR values, which are known to grow in tandem with JIF, are amplified in such a way that they track the latter metric more closely than in the quadratic transformation.

While the cubic version of the JIF metric (Cubic JIF) also exhibits high correlation values with the SJR based metrics, these values are comparatively lower overall. For Pearson correlations, this outcome is expected, as cubic transformations tend to exaggerate differences and thereby amplify disparities in absolute values. The trends observed for the Quadratic JIF metric are mirrored in it's cubic counterpart with respect to the Spearman correlation: a higher value is recorded between Metrics 2 and 5, while a lower value is observed between Metrics 1 and 5. Notably, however, the Spearman coefficient between Metrics 2 and 4 exceeds that between Metrics 2 and 5. This suggests that a balance was effectively achieved by combining the cubic transformation with SJR and the quadratic transformation with JIF, yielding the highest correlation coefficient observed between these two indicators.

In summary, the cubic transformations of the proposed metrics produced higher Spearman correlation values. This result is coherent with the intended purpose of the cubic

transformations, namely to further accentuate the magnitude of metric values and thereby reduce the relative weight of individual citations in the calculation. Consequently, the pairs of metrics with the strongest correlation, SJR and JIF, exhibit even stronger correlations in their cubic-transformed versions. For the purposes of this study, this implies that researchers ranking closely under Metrics 1 or 2, or Metrics 4 or 5, may find it challenging to distinguish between these proposed metrics, given their similarity. Tying back to the respondent's answer mentioned in the beginning of this section, they found that their sorting of the publications was closest to that of the Quadratic JIF's, however, they noted that both the Quadratic SJR and Cubic JIF metrics were very similar, just like it was predicted with these correlations.

In contrast to these conclusions, only a few notable or high correlation values are observed in comparisons involving H-index–based proposed metrics. This finding further reinforces the earlier observation that the H-index exhibits trends and behavior distinct from those of the other four proposed metrics.

An additional observation concerns the H-Index values, where one of the most frequent values identified was 446. This figure is notable for being the highest H-Index across all citations, and by a considerable margin. What made this observation particularly striking was that the corresponding SJR and JIF values were consistently low. This anomaly led to the identification of the journal in question: Lecture Notes in Computer Science (LNCS). This finding initially caused confusion, as LNCS is a highly recognized and influential book series/journal in the field of Computer Science.

The apparent contradiction became clear upon further investigation. Data from the SCImago Journal and Country Rank platform revealed that LNCS has historically published an extraordinary volume of documents. Over the past 26 years, LNCS published on average 19,669.08 citable and non-citable documents annually, with an average of 19,417.33 documents over the last three years. At the same time, the average cites per document over a three-year window was 0.952 across 26 years, with a slightly higher value of 1.463 in 2024. This situation can be summarized as follows: LNCS is a long-established and highly prolific venue, but its publications are not frequently cited.

This pattern is reflected in the metrics in an entirely predictable way, consistent with the principles outlined in the literature review. LNCS achieves a very high H-Index because of its immense productivity. Across many years and thousands of publications, some works inevitably accumulate substantial citations, increasing the H-Index over time. Importantly, the H-Index does not decrease, even if many other works have limited impact. Thus, the journal accrues a high H-Index despite most of its individual publications lacking significant influence or prestige.

By contrast, both SJR and JIF assign LNCS very low values. In the case of SJR, this outcome reflects the journal's limited prestige: its massive volume of publications yields relatively few citations per document, and when citations do occur, they are often from journals of similarly modest impact. The JIF, being sensitive to citation averages over

a short time window, is also low for the same reasons. Moreover, JIF values tend to be distorted by a small number of highly cited papers in otherwise low-impact venues. For LNCS, however, the overwhelming number of lowly cited publications dilutes the effect of any highly cited outliers, keeping the average low.

While one could speculate about the influence of citation malpractice on these metrics, the enormous scale of LNCS publication output makes it difficult to determine whether malpractice plays a significant role. The dynamics of the H-Index in this context are illustrated in Figure 2 of Gómez-Núñez et al. [**19**], where a scientogram shows how journals in the SJR database are clustered according to citation, co-citation, and coupling patterns. Within this visualization, LNCS appears as the largest journal in the field of Computer Science. Its vast publication history results in few citations per document but a very large aggregate citation count, creating a divergence between metrics. Thus, SJR and JIF capture LNCS as a venue of relatively low prestige and impact, whereas the H-Index portrays it as extremely productive and influential over time. Once again, this distinction coincides with the interpretations of metric behavior discussed in the related work.

CHAPTER 5

Survey Answers

The complete survey that was presented to the researchers is detailed in Appendix A, Researcher Survey. This section shows how the data was presented, introduced, and how the survey itself was conducted.

A total of 13 responses were gathered from the survey. Table 5.1 summarizes the choice of each one of the metrics, answering the questions in the survey shown in A.2 where the most popular metric turned out to be the Quadratic JIF, with 5 respondents choosing it as the metric which best reflects their rankings. This is not too surprising, given that JIF, as mentioned before, is one of the most prominent Journal-Level metrics, and the quadratic and cubic versions of the metrics favor higher JIF values. The survey results show that researchers also tend to recognize JIF as the most relevant metric. With three respondents quoting it as the metric that they immediately chose to use. Besides this, researchers seemed split on which metric to choose. One respondent found that no metric truly reflected their interpretation of a publications' impact since this can only be understood by analyzing the work itself and not via any quantitative methods and refused to choose one metric. That is a highly valid criticism, yet one that goes against the goals and scope of this work, as will be discussed in Chapter 6. Two researchers selected the SJR-Adjusted Citation impact metrics as their most representative metric, Quadratic H-Index was chosen twice and three other researchers chose Cubic JIF.

TABLE 5.1. Frequency of Metric Selection by Researchers

	Quadratic SJR	Cubic SJR	Quadratic H-Index	Quadratic JIF	Cubic JIF
Times Chosen	2	0	2	5	3

Four submissions provided alternative rankings together with their answers.

One of the respondents who chose Quadratic SJR noted that the only difference between the Quadratic SJR ranking and their own proposed ranking was the order of the top two publications. According to this researcher, Publication 9 was the most impactful, surpassing Publication 1, which was considered the best by the metric. This was because its Field Citation Ratio[1] was 10, meaning that it has been cited ten times more than the average article published in the same year and within the same field(s). They also mentioned that while they would place Publication 9 in first place, they admitted that metrics 1, 2 and 4 already rank this publication very highly, as the second most impactful publication. This researcher's ranking clearly demonstrated that, by introducing a slight

[1]Field Citation Ratio's calculation

variation to the ordering produced by Quadratic SJR, the proposed metric was able to closely reflect the researcher's preferences. It is possible that incorporating adjustments, such as assigning different weights to more influential publications within their respective fields, could further enhance the accuracy of these metrics. This approach could be further extended by assigning greater weight to citations originating from the same field as the cited publication, with even greater weights applied depending on the relative influence of those citations within the field.

The rest of the researchers found that the order of the metric they chose reflected their own, saying that they would not change any publication in their ranking. Three other respondents were of note because they submitted their ordering in their entirety.

TABLE 5.2. Researcher and Quadratic and Cubic JIF Rankings

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
Respondent 1	9	1	8	17	2	18	16	19	4	15	3	7	5	12	13	11	20	14	6	10
Respondent 2	17	20	13	7	4	9	8	6	5	3	12	1	18	14	16	2	11	19	10	
Respondent 3	3	8	16	4	19	17	18	15	14	20	9	13	2	1	7	12	5	10	11	6
Quadratic JIF	1	9	2	8	11	3	5	17	4	6	15	12	19	7	18	20	16	13	14	10
Cubic JIF	1	3	9	8	5	2	15	18	17	12	14	11	4	13	7	19	16	6	20	10

Table 5.2 presents the rankings of the researchers who selected the Quadratic JIF metric as their most representative and respondent 3's ranking who chose the Cubic JIF's metric. These are then followed by the rankings of these metrics. Based solely on these tables, it is challenging to assess the degree of agreement between these rankings. However, alternative approaches exist to quantify the proximity of these rankings without relying exclusively on the tables.

Kendall's Tau can be employed to evaluate the degree of concordance between the respondents' rankings and those of the metrics [**20**]. Additionally, Rank Bias Overlap (RBO) can be used to measure the extent of overlap between the rankings while placing greater emphasis on the top-ranked items [**47**]. This weighting is particularly relevant in the present context, as researchers may have devoted more attention and detail to the initial positions in their rankings, rendering the initial ranks more meaningful than the later ones. Additionally, RBO was used due to a missing value in Answer 2, where publication 15 was omitted, likely inadvertently, from the respondent's ranking. This distinction is important because Kendall's Tau requires that the paired lists have identical lengths, whereas RBO does not; it evaluates the overlap at varying depths of the lists. The following values were computed using the `kendalltau` function from the `scipy.stats` library and the separate RBO library in Python.

Figure 5.1 presents the Kendall Tau correlation analysis for the rankings proposed by respondents 1 and 3 against the rankings of each one of the 5 metrics. For Respondent 1, the metric with the highest correlation to their ranking was Quadratic JIF, with a coefficient of 0.589, indicating a moderate level of concordance. Notably, Respondent 1 provided details regarding the methodology they employed to rank the publications, which involved considering the average number of citations per year and the total number of citations, but only including citations originating from journals with a JIF of 3 or

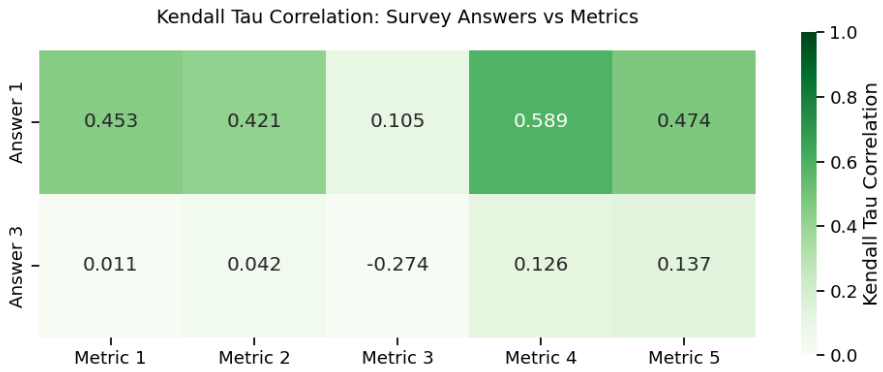


FIGURE 5.1. Correlation between Answers and Metrics

higher. Given this approach, the fact that the moderate correlations with Metrics 4 and 5 are not of a higher value is understandable, as the researcher's own ranking mechanism is based on different criteria. In contrast, for Respondent 3, the correlation coefficients are all weak or negligible. Notably, the highest correlation is a negative value observed between this ranking and Quadratic H-Index, indicating that the respondent's ranking contradicts the ordering produced by the quadratic formulation of the H-Index. The highest positive correlation is observed with the Cubic JIF metric, which is consistent with the respondent's choice. While these correlation values are weak, they remain informative. The low coefficients suggest that Respondent 3 largely relied on their own criteria to generate their rankings. This is corroborated by the respondent's explanation that their ranking methodology combined JIF values, citation counts, and the age of the publication. These are factors which are not fully captured by the metrics under study.

It is important to note that, despite the overall low correlation values shown in Figure 5.1, minor differences among the metric coefficients can result in meaningful differences in the ordering of publications. Even seemingly small shifts in rank, for instance, a publication being placed one or two positions higher or lower, can significantly influence a researcher's judgment regarding which metric best reflects their ranking. Consequently, these small variations in low correlation coefficients remain relevant in determining the chosen metric. In other words, although there is minimal correlation between Respondent 3's ranking and Cubic JIF, this metric nonetheless best reflects the researcher's ranking in comparison to the other values. Furthermore, it is likely that the small differences in ranking reflected by the correlation coefficients were concentrated on a subset of publications, which could have been the primary focus of the researcher, perhaps due to the highly polarizing nature or exceptional impact of these publications, for example. Consequently, researchers may assign greater significance to the relative ordering of this core subset than to the overall ranking. RBO is particularly suited to this context, as it measures the similarity between rankings while assigning greater weight to the initial positions, with the bias decreasing over the length of the ranking. Since researchers tend to focus on the most polarizing or impactful publications, RBO provides a meaningful assessment of the degree of overlap while emphasizing the perceived importance of top-ranked items. Moreover,

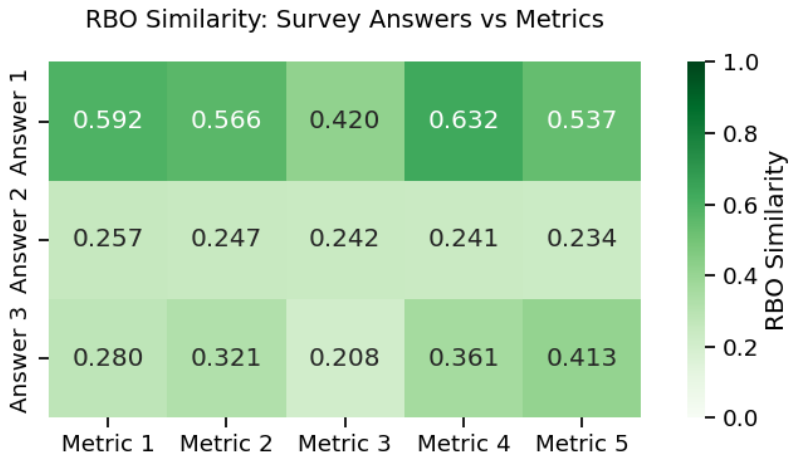


FIGURE 5.2. RBO Similarities

RBO allows the inclusion of Respondent 2, whose ranking contained a missing value, as it does not require that lists have identical lengths. Figure 5.2 evaluates the RBO similarity across the rankings of all respondents, taking into account the emphasis on the most influential publications. For example, the top 10 publications carry about 86% of the weight of the final RBO value [**47**]. It is important to emphasize that RBO is not a correlation coefficient like Kendall's Tau; consequently, the values should not be compared directly with one another and cannot be interpreted in the same manner. These values indicate that, for Respondent 1, their ranking exhibits the greatest overlap with Quadratic JIF, reflecting what was found with the Kendall Tau values. Respondent 2's ranking demonstrates no significant overlap with any metric. The highest overlap was observed with the Quadratic SJR. However, all values are relatively similar. This respondent noted that their ranking contained seven publications similarly ranked with Quadratic JIF and six publications similarly ranked with Quadratic SJR and Cubic JIF. This negligible difference across metrics is effectively what is being captured by RBO. Finally, the Respondent 3 RBO values corroborate the findings from Kendall's Tau. While all values are generally low to moderate, there is an increase in overlap with Cubic JIF. This indicates that, when prioritizing the top-ranked publications in the weighting of overlaps, Cubic JIF continues to emerge as the metric most closely aligned with the researcher's ranking.

What we find is, once again, low to moderate values across all comparisons. However, there is an increase of Quadratic JIF for respondents 1 and 2, once again reinforcing the idea that there are slight differences in these orderings made by the researchers that led them to choosing the quadratic JIF-Adjusted Citation Impact metric. This is especially apparent with the ideas mentioned by respondent 2 who said that there were 7 publications which they found were close between their ranking and Quadratic JIF. All respondents also found that Quadratic SJR, Quadratic JIF and Cubic JIF were quite similar to in rankings overall. Finally, most respondents found that Cubic SJR was the least representative

metric. This is reflected in the overall survey answers where no researcher chose Cubic SJR.

Based on these correlation values, several conclusions can be drawn. Within our sample of researchers, Quadratic JIF emerged as the most frequently selected, followed by Metrics 1 and 5. Additionally, among the respondents who provided their rankings, although Quadratic JIF was chosen as the metric that best reflected their assessments, the observed correlations or similarities with the proposed metrics were generally weak to moderate. This indicates that, while Quadratic JIF, or its cubic counterpart in the instance of respondent 3, was the metric most closely aligned with their rankings, these metrics did not fully capture the experts' underlying methods for evaluating and ranking publications. Showing that further improvements to them will still be needed.

CHAPTER 6

Conclusion

Based on the findings obtained from the survey responses, the quadratic JIF-Adjusted Citation Impact metric, emerged as the most frequently selected proposed metric. However, among the respondents who provided their rankings and chose Quadratic JIF, the correlation between their rankings and those produced by the metrics was found to overall be weak to moderate. This indicates that, although Quadratic JIF was the most representative of these specific researchers' rankings, it did not fully reflect the methods they employed to rank the publications. Consequently, while it is the most popular, in some cases it is insufficient to capture the rankings produced by researchers, suggesting that further adjustments are necessary to fine-tune the metric to accurately reflect expert assessments.

Additionally, the Cubic JIF metric was identified as the second most popular. This finding is somewhat expected, as the Quadratic transformation of this metric was the most popular, and the different transformations of these two metrics do not introduce substantial differences in their rankings, given their very strong positive correlations. These were followed in popularity by the quadratic SJR-Adjusted Citation Impact Metric and its H-Index based counterpart. Some researchers reported that, in cases where JIF-based metrics were selected as the most representative, the rankings generated by the Quadratic SJR closely aligned with their own rankings, consistent with the observed Pearson and Spearman correlation results. Therefore, although Quadratic JIF emerged as the most popular, it is plausible that researchers found it challenging to distinguish between the Quadratic and Cubic JIF-Adjusted Citation Impact metrics and the Quadratic SJR metric, as these demonstrate similar behaviors and trends, analogous to the raw indicators on which they are based.

Thus, two perspectives emerge. On the one hand, the Quadratic JIF appears to be the metric that best reflects expert rankings. On the other hand, the Quadratic SJR, although less capable of directly capturing these rankings, is based on the SJR indicator, which is recognized as more robust and less susceptible to citation manipulation techniques. Furthermore, since the Quadratic SJR correlates strongly with the Quadratic JIF, the differences in rankings produced by these two metrics are likely to be minimal. As a result, in cases where Quadratic JIF closely reflects expert rankings, it is reasonable to expect that the quadratic SJR-Adjusted Citation Impact metric produces a similar ordering, with the added advantage of being grounded in a more reliable and well-rounded indicator.

6.1. Limitations and Future Work

While our proposed metrics allow for a new measure of the impact of a publication based on the impact of its citing articles' journals, there are some aspects to this study that have limited the extent to which our metric is able to be calibrated to be the truest possible interpretation of impact.

A key factor to consider when interpreting the findings of this study is the number of survey responses received. Due to time constraints, a total of 13 responses were collected from researchers affiliated with various institutions, departments, and research units. All respondents had a minimum of five years of research experience, with the majority having more than ten years of experience. In future studies, expanding the sample size would provide a more robust representation of researchers' rankings and offer a clearer understanding of the most widely supported metric. Increasing the number of responses, particularly those including experts' full rankings, would enable a more detailed comparison of how these rankings align with the metrics' outputs, thereby offering stronger guidance on how to refine the most promising proposed metric.

Another aspect worth considering involves the number and variety of publications included in the study. For this analysis, 20 publications were gathered for use in the rankings. Although these publications covered diverse fields and subjects, the majority were concentrated in the field of computer science. Expanding the dataset to include a broader and slightly larger set of publications could reveal useful inconsistencies in the rankings produced by the proposed metrics. This, in turn, would allow future surveys to provide a more comprehensive understanding of how the metrics behave, thereby supporting a more thorough and fairer interpretation of their performance.

Given the scope of our work, some important aspects of impact evaluation were sidelined. Our primary aim was to assess the impact of a publication leveraging journal-level metrics of the citing articles. As a result, variations in citation patterns across different academic fields were not taken into account. While metrics such as SNIP (Source Normalized Impact per Paper) and Elsevier's Field-Weighted Citation Impact provide some valuable insights into the impact of a publication within its specific field of study, we deliberately chose not to incorporate these metrics into our analysis. This decision, though consistent with the focus of our work, has been recognized as a point of frustration by some respondents in our survey. Several participants expressed that, with the lack of field-specific adjustments in our citation impact calculations, they were unable to rely on one metric to encapsulate their ranking. In hindsight, it may be more appropriate to combine a metric like SNIP with the SJR index, as this could yield a more nuanced understanding of citation impact. Alternatively, prioritizing citations from sources marked as belonging to the same or closely related fields could offer a more field-biased measure of impact.

The formulation of the proposed metrics incorporates the concept of adjusting the importance of a publication's journal, represented by the symbol λ. However, in the context of this study, the value of λ was consistently set to 1, thereby assigning the

publication's own journal the same weight as a citation. This approach may be perceived by some researchers as an unrealistic interpretation of the publication's impact. Perhaps for future expansions on this work, more weight could be placed on the publications' journal metrics.

While the SJR metric restricts self-citations, the proposed metrics in this study do not. At present, both citing articles and self-citations are assigned equal weight, a practice that has been noted by respondents as a point for potential modification. Some participants suggested that a factor could be introduced to reduce the impact of self-citations in the calculation, thereby providing a more accurate representation of a publication's true citation impact.

If the JIF continues to demonstrate its efficacy in capturing researchers' impact assessments, it may be insightful to compare it with the JIF data excluding self-citations, as provided by Clarivate. Such a comparison could potentially bring JIF-based metrics more in line with some researchers' rankings, while also enhancing the robustness of JIF, bringing it closer to the SJR in terms of analytical depth. However, as discussed, the complete removal of self-citations in these alternative metrics may be viewed as excessive and could undermine certain natural aspects of scientific research.

Finally, as suggested by some respondents in the survey, incorporating the age of an article into the calculation of its impact could provide valuable insights. Specifically, an article that was highly impactful in the past, when its topic was particularly relevant, may no longer hold the same weight in terms of impact as a more recent publication. Including a time-sensitive element, akin to the approach used in metrics such as the SJR or JIF, could offer a more comprehensive interpretation of impact that aligns more closely with researchers' assessments.

6.2. Closing Remarks

The conclusions presented at the beginning of this chapter address Research Question 3. The quadratic JIF-Adjusted Citation Impact proposed metric demonstrates the closest alignment with researcher assessment. Section 6.1 addresses the latter part of this question, exploring how further research in this area could lead to adjustments in the proposed metrics to better align with researchers' expectations of impact. Potential directions for future research include the incorporation of limitations on self-citations within the proposed metrics, the introduction of citation value weighting based on the age of the publication and the consideration of differentiated weighting for citations originating from the same journal as the publication.

While the Quadratic JIF emerges as the most widely favored metric among researchers, our SJR-adjusted Citation Impact relies on a more robust framework that mitigates the influence of citation manipulation tactics. Moreover, it has been demonstrated that both of these metrics have strong correlations, indicating that the final ranking differences between these proposed metrics are likely to be minimal. This suggests that, although

further research is necessary to identify the optimal metric that balances researcher feedback, prevents citation manipulation, and accurately reflects the transfer of impact from citations to publications, this study provides a solid foundation for identifying the most critical metrics and outlining potential directions for future work. Ultimately, these efforts aim to establish a metric capable of more accurately assessing the true impact of citations on scientific publications.

6.3. Communication of Ethics Committee Approval

Before presenting the survey to any researcher, approval was requested from the ISTA Ethics Committee of ISCTE-IUL to authorize the conduct of the interviews. The Committee's approval of the request is shown in Appendix B, Ethics Committee Acknowledgment.

6.4. Communication of Research Dissemination

During the course of this study, a Research in Progress article was accepted as a poster for presentation at the 25th Conference of the Portuguese Association for Information Systems (CAPSI 2025), which will take place in the Azores. As of the date of writing, September 29, 2025, the article has been reviewed, accepted, and is currently awaiting presentation, scheduled for October 16-18, 2025.

Bibliography

[1] Ashok Agarwal, Damayanthi Durairajanayagam, Sindhuja Tatagari, SandroC Esteves, Avi Harlev, Ralf Henkel, Shubhadeep Roychoudhury, Sheryl Homa, NicolásGarrido Puchalt, Ranjith Ramasamy, Ahmad Majzoub, KimDao Ly, Eva Tvrda, Mourad Assidi, Kavindra Kesari, Reecha Sharma, Saleem Banihani, Edmund Ko, Muhammad Abu-Elmagd, Jaime Gosalvez, and Asher Bashiri. Bibliometrics: tracking research impact by selecting the appropriate metrics. *Asian Journal of Andrology*, 18(2):296, 2016.

[2] Manolis Antonoyiannakis. Impact factor volatility due to a single paper: A comprehensive analysis. *Quantitative Science Studies*, 1(2):639–663, June 2020.

[3] Clas Bergstrom. Eigenfactor: Measuring the value and prestige of scholarly journals. *College and Research Libraries News*, 68:314–316, 05 2007.

[4] Henry Bi. Four problems of the h-index for assessing the research productivity and impact of individual authors. *Scientometrics*, 128, 03 2022.

[5] Mario Biagioli. Watch out for cheats in citation game. *Nature*, 535(7611):201–201, July 2016.

[6] Johan Bollen, Marko A. Rodriquez, and Herbert Van De Sompel. Journal status. *Scientometrics*, 69(3):669–687, December 2006.

[7] Lisa Colledge, Félix De Moya-Anegón, Vicente Guerrero-Bote, Carmen López-Illescas, M'hamed El Aisati, and Henk Moed. SJR and SNIP: two new journal metrics in Elsevier's Scopus. *Serials: The Journal for the Serials Community*, 23(3):215–221, January 2010.

[8] I. Diane Cooper. Bibliometrics basics. *Journal of the Medical Library Association : JMLA*, 103(4):217–218, October 2015.

[9] Jingda Ding, Chao Liu, and Goodluck Asobenie Kandonga. Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. *Scientometrics*, 122(3):1303–1322, March 2020.

[10] Iztok Fister jr, Iztok Fister, and Matjaž Perc. Toward the discovery of citation cartels in citation networks. *Frontiers in Physics*, 4, 12 2016.

[11] Eric A. Fong and Allen W. Wilhite. Authorship and citation manipulation in academic research. *PLOS ONE*, 12(12):1–34, 12 2017.

[12] Grischa Fraumann, Jennifer D'Souza, and Kim Holmberg. 4.7 Eigenfactor. In Rafael Ball, editor, *Handbook Bibliometrics*, pages 245–254. De Gruyter, January 2021.

[13] C. García-Villar and J.M. García-Santos. Indicadores bibliométricos para evaluar la actividad científica. *Radiología*, 63(3):228–235, May 2021. Publisher: Elsevier BV.

[14] Grey Giddins. Journal metrics. *Journal of Hand Surgery (European Volume)*, 39(8):807–807, 2014. PMID: 25236532.

[15] Borja González-Pereira, Vicente P. Guerrero-Bote, and Félix Moya-Anegón. A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*, 4(3):379–391, July 2010.

[16] Victor Grech. Publish or perish, information overload, and journal impact factors – A conflicting tripod of forces. *Saudi Journal of Anaesthesia*, 16(2):204–207, April 2022.

[17] Thorsten Gruber. Academic sell-out: how an obsession with metrics and rankings is damaging academia. *Journal of Marketing for Higher Education*, 24(2):165–177, July 2014.

[18] Vicente P. Guerrero-Bote and Félix Moya-Anegón. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, 6(4):674–688, October 2012. Publisher: Elsevier BV.

[19] Antonio J. Gómez-Núñez, Benjamin Vargas-Quesada, Zaida Chinchilla-Rodríguez, Vladimir Batagelj, and Félix Moya-Anegón. Visualization and analysis of SCImago Journal & Country Rank structure via journal clustering. *Aslib Journal of Information Management*, 68(5):607–627, September 2016. Publisher: Emerald.

[20] K. H. Hamed. The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data. *Hydrological Sciences Journal*, 56(5):841–853, July 2011.

[21] Petr Heneberg. From excessive journal self-cites to citation stacking: Analysis of journal self-citation kinetics in search for journals, which boost their scientometric indicators. *PLOS ONE*, 11(4):1–20, 04 2016.

[22] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.

[23] John P. A. Ioannidis, Kevin Boyack, and Paul F. Wouters. Citation metrics: A primer on how (not) to normalize. *PLOS Biology*, 14(9):1–7, 09 2016.

[24] John PA Ioannidis and Kevin W Boyack. Citation metrics for appraising scientists: misuse, gaming and proper use. *Medical Journal of Australia*, 212(6):247–249.e1, 2020.

[25] Kihong Kim and Yeonok Chung. Overview of journal metrics. *Science Editing*, 5(1):16–20, February 2018. Publisher: Korean Council of Science Editors.

[26] Frank-Thorsten Krell. The Journal Impact Factor as a performance indicator. *European Science Editing*, 38:3–6, February 2012.

[27] Loet Leydesdorff, Paul Wouters, and Lutz Bornmann. Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators—a state-of-the-art report. *Scientometrics*, 109(3):2129–2150, December 2016. Publisher: Springer Science and Business Media LLC.

[28] Nan Ma, Jiancheng Guan, and Yi Zhao. Bringing PageRank to the citation analysis. *Information Processing & Management*, 44(2):800–810, March 2008.

[29] Jorge Mañana-Rodríguez. A critical review of SCImago Journal & Country Rank. *Research Evaluation*, 24(4):343–354, October 2015.

[30] John Mingers and Liying Yang. Evaluating journal quality: A review of journal citation indicators and ranking in business and management. *European Journal of Operational Research*, 257(1):323–337, February 2017.

[31] Durga Prasanna Misra. Publication Metrics: Types, Utility, and Caveats. *Indian Journal of Rheumatology*, 17(2S), September 2022.

[32] Henk Moed. Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4:265–277, 11 2009.

[33] Himel Mondal and Shaikat Mondal. A brief review on article-, author-, and journal-level scientometric indices. *Indian Dermatology Online Journal*, 13(5):578, 2022.

[34] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking : Bringing Order to the Web. In *The Web Conference*, 1999.

[35] Upeksha Perera and Manjula Wijewickrema. Relationship between Journal-Ranking Metrics for a Multidisciplinary Set of Journals. *portal: Libraries and the Academy*, 18(1):35–58, 2018.

[36] S.U. Pillai, T. Suel, and Seunghun Cha. The Perron-Frobenius theorem: some of its applications. *IEEE Signal Processing Magazine*, 22(2):62–75, March 2005.

[37] D. Rodrigues, A.L. Lopes, and F. Batista. Web of Science Citation Gaps: An Automatic Approach to Detect Indexed but Missing Citations. Master's thesis, Iscte - University Institute of Lisbon, 2023.

[38] Ernesto Roldan-Valadez, Shirley Salazar-Ruiz, Rafael Ibarra, and Camilo Rios. Current concepts on bibliometrics: a brief review about impact factor, eigenfactor score, citescore, scimago journal rank, source-normalised impact per paper, h-index, and alternative metrics. *Irish Journal of Medical Science (1971 -)*, 188, 12 2018.

[39] Patrick Schober, Christa Boer, and Lothar A. Schwarte. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, May 2018.

[40] U. Senanayake, M. Piraveenan, and A.Y. Zomaya. The p-index: Ranking Scientists Using Network Dynamics. *Procedia Computer Science*, 29:465–477, 2014.

[41] Angelo Solimini and Renata Solimini. Impact factor and other metrics for evaluating science: Essentials for public health practitioners. *Italian Journal of Public Health*, 8:96–103, 03 2011.

[42] Iman Tahamtan and Lutz Bornmann. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics*, 121(3):1635–1684, December 2019.

[43] Chandrasegar Thirumalai, Swapna Anupriya Chandhini, and M Vaishnavi. Analysing the concrete compressive strength using Pearson and Spearman. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, pages 215–218, Coimbatore, April 2017. IEEE.

[44] Daniel Torres-Salinas and Evaristo Jiménez-Contreras. Introducción y estudio comparativo de los nuevos indicadores de citación sobre revistas científicas en *Journal Citation Reports* y *Scopus*. *El Profesional de la Informacion*, 19(2):201–208, March 2010.

[45] Ludo Waltman, Nees Jan Van Eck, and Paul Wouters. Counting publications and citations: Is more always better? *Journal of Informetrics*, 7(3):635–641, July 2013.

[46] Mingyang Wang, Jinjin Ren, Shi Li, and Guangsheng Chen. Quantifying a Paper's Academic Impact by Distinguishing the Unequal Intensities and Contributions of Citations. *IEEE Access*, 7:96198–96214, 2019.

[47] William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4):1–38, November 2010.

[48] Jevin D. West, Theodore C. Bergstrom, and Carl T. Bergstrom. The Eigenfactor MetricsTM: A Network Approach to Assessing Scholarly Journals. *College & Research Libraries*, 71(3):236–244, May 2010.

[49] Jason Yuen. Comparison of Impact Factor, Eigenfactor Metrics, and SCImago Journal Rank Indicator and h-index for Neurosurgical and Spinal Surgical Journals. *World Neurosurgery*, 119:e328–e337, November 2018. Publisher: Elsevier BV.

APPENDIX A

Researcher Survey

A.1. Citation Table

Before the survey was presented to the researchers, they were given a set of 20 publications displayed in the table shown in Figure 4.4 in Section 4.2. What follows is the explanation these researchers were given about this table and the objective of their task.

> The table below shows, for each one of the publications, the SCImago Journal Rank (SJR), H-Index and Journal Impact Factor (JIF) of each one of its citations.
>
> By clicking the ⓘ of the numbered publications or the citations, you will be sent to the website related to the DOI.
>
> You can sort the publications based the publication year, number of citations column and the 5 predefined metrics (metric 1.. metric 5).
>
> Please select the metric that better sorts the publications, based on criteria you think should be the best.
>
> After sorting, if you still think that the results could be improved even further, please report the final most suitable order, by descending relevance (ex: publication 1, 3, 5, 7, 8...)
>
> Click here to access the survey to submit your answer
>
> By clicking the headers of the metrics columns, values of the table are sorted by that column.

After this introduction was made, the table was presented as shown in Figure A.1. Below the table, the proposed metrics were shown, as seen in Figure A.2

A.2. Researcher Survey

After interpreting the table and determining which ranking of publications best corresponded to each researcher's own assessment, participants proceeded to the survey, which was split into 4 sections, where they were asked the following questions:

- Consent was requested as shown in Figure A.3.
- Which metric best reflected the Researcher's ranking, as seen in Figure A.4. Additionally, they were asked:
  - If the Researcher had a key difference between the ranking of the metric and their own. (Optional)
  - What was their ranking of the publications. (Optional)
  - Any further suggestions were requested. (Optional)

The table below shows, for each one of the publications, the SCImago Journal Rank (SJR), H-Index and Journal Impact Factor (JIF) of each one of its citations.

By clicking the ⓘ of the numbered publications or the citations, you will be sent to the website related to the DOI.

You can sort the publications based the publication year, number of citations column and the 5 predefined metrics (metric 1.. metric 5).

Please select the metric that better sorts the publications, based on criteria you think should be the best.

After sorting, if you still think that the results could be improved even further, please report the final most suitable order, by descending relevance (ex: publication 1, 3, 5, 7, 8…)

Click here to access the survey to submit your answer

By clicking the headers of the metrics columns, values of the table are sorted by that column.

SJR/H-Index/JIF values for each Cited Pub DOI

								Publication				Citation 1				Citation 2				Citation 3				Citation 4				Citation 5			
DOI	Pub Year	#Citations	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5	SJR	H	JIF	ⓘ	SJR	H	JIF	ⓘ	SJR	H	JIF	ⓘ	SJR	H	JIF	ⓘ	SJR	H	JIF	ⓘ	SJR	H	JIF	ⓘ
Publication 1 ⓘ	2017	29	30.372	21.975	763.231	48.889	31.397	3.154	205	4.990	ⓘ	14.605	109	14.900	ⓘ	14.605	109	14.900	ⓘ	10.486	306	17.090	ⓘ	7.706	75	16.480	ⓘ	4.321	142	12.230	ⓘ
Publication 2 ⓘ	2018	37	11.968	7.456	509.679	30.894	20.278	0.740	30	2.190	ⓘ	3.403	159	8.860	ⓘ	3.383	236	12.480	ⓘ	3.068	236	8.680	ⓘ	2.228	201	4.920	ⓘ	2.217	151	7.020	ⓘ
Publication 3 ⓘ	2023	5	9.562	8.031	344.652	32.474	28.181	1.749	137	9.220	ⓘ	6.280	213	25.700	ⓘ	3.382	179	9.220	ⓘ	1.749	137	9.220	ⓘ	0.763	49	4.230	ⓘ	0.000	0	0.000	ⓘ
Publication 4 ⓘ	2021	26	8.886	5.434	563.670	22.089	14.275	1.008	45	7.340	ⓘ	2.643	161	6.650	ⓘ	1.091	193	4.550	ⓘ	1.091	193	4.550	ⓘ	1.091	193	4.550	ⓘ	0.995	33	6.610	ⓘ
Publication 5 ⓘ	2016	19	9.311	6.567	524.278	28.963	22.458	0.729	170	3.160	ⓘ	4.557	136	19.120	ⓘ	2.070	249	9.430	ⓘ	1.875	271	9.290	ⓘ	0.848	170	4.270	ⓘ	0.745	65	4.930	ⓘ

FIGURE A.1. Page Presented to Researchers

Legend: Metrics

- Metric 1: $\sqrt[3]{\lambda(SJR_p+1)^2+\sum(SJR_i+1)^2}$
- Metric 2: $\sqrt[3]{\lambda(SJR_p+1)^3+\sum(SJR_i+1)^3}$
- Metric 3: $\sqrt[3]{\lambda(HIndex_p+1)^2+\sum(HIndex_i+1)^2}$
- Metric 4: $\sqrt[3]{\lambda(JIF_p+1)^2+\sum(JIF_i+1)^2}$
- Metric 5: $\sqrt[3]{\lambda(JIF_p+1)^3+\sum(JIF_i+1)^3}$

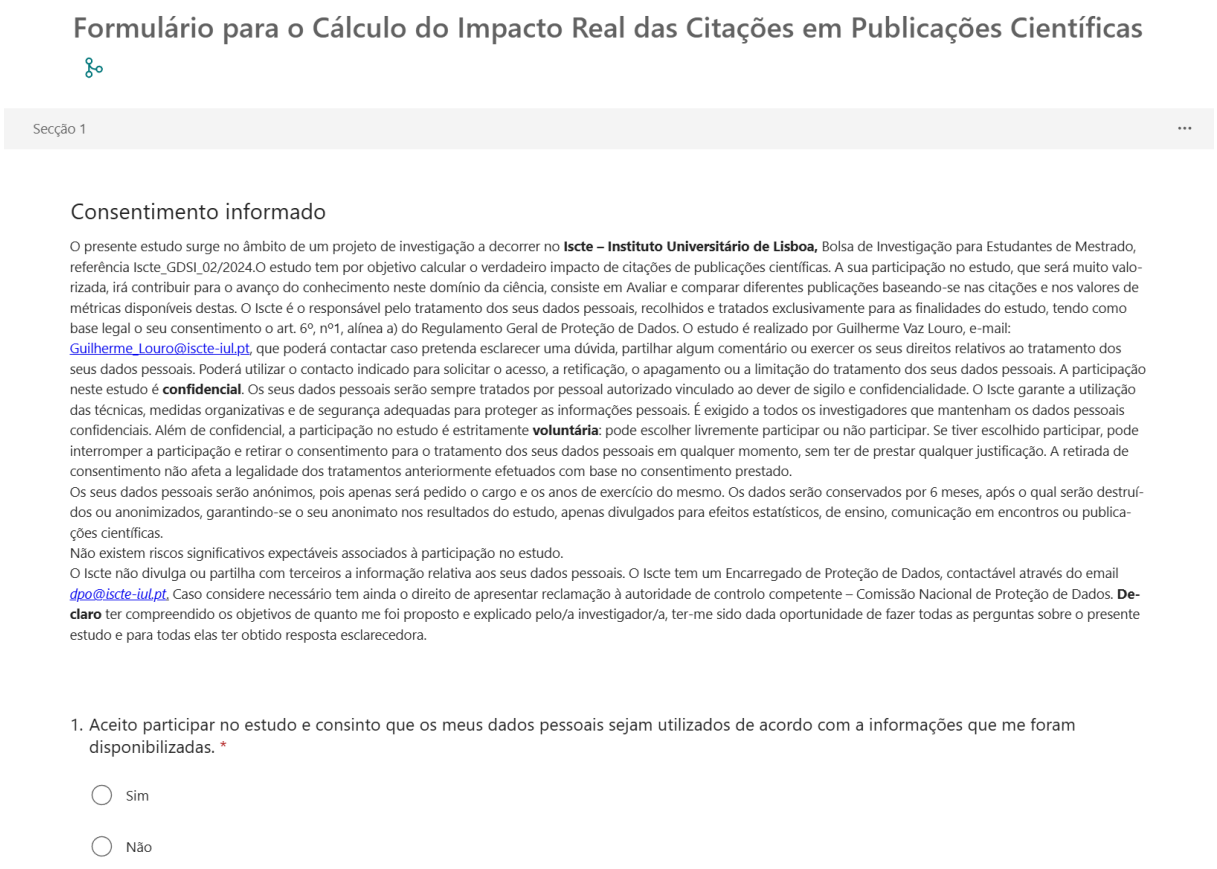
Where: SJR_p is the SJR of the publication's journal and SJR_i is the SJR of a citing publications journal. The same is repeated for the other metrics.

λ is a weighting factor that can be adjusted based on the importance of the publication. Currently λ is set to 1.

FIGURE A.2. Proposed Metrics

- Job, Department, Research Unit and years of experience in research were asked as shown in Figure A.5. All of these question were optional.
- The respondent was thanked for their participation, as shown in Figure A.6.

where they were asked to indicate which metric best reflected this ranking and, if they wished, to submit their own ranking along with any additional suggestions.



Formulário para o Cálculo do Impacto Real das Citações em Publicações Científicas

Secção 1

Consentimento informado

O presente estudo surge no âmbito de um projeto de investigação a decorrer no **Iscte – Instituto Universitário de Lisboa,** Bolsa de Investigação para Estudantes de Mestrado, referência Iscte_GDSI_02/2024.O estudo tem por objetivo calcular o verdadeiro impacto de citações de publicações científicas. A sua participação no estudo, que será muito valorizada, irá contribuir para o avanço do conhecimento neste domínio da ciência, consiste em Avaliar e comparar diferentes publicações baseando-se nas citações e nos valores de métricas disponíveis destas. O Iscte é o responsável pelo tratamento dos seus dados pessoais, recolhidos e tratados exclusivamente para as finalidades do estudo, tendo como base legal o seu consentimento o art. 6º, nº1, alínea a) do Regulamento Geral de Proteção de Dados. O estudo é realizado por Guilherme Vaz Louro, e-mail: Guilherme_Louro@iscte-iul.pt, que poderá contactar caso pretenda esclarecer uma dúvida, partilhar algum comentário ou exercer os seus direitos relativos ao tratamento dos seus dados pessoais. Poderá utilizar o contacto indicado para solicitar o acesso, a retificação, o apagamento ou a limitação do tratamento dos seus dados pessoais. A participação neste estudo é **confidencial**. Os seus dados pessoais serão sempre tratados por pessoal autorizado vinculado ao dever de sigilo e confidencialidade. O Iscte garante a utilização das técnicas, medidas organizativas e de segurança adequadas para proteger as informações pessoais. É exigido a todos os investigadores que mantenham os dados pessoais confidenciais. Além de confidencial, a participação no estudo é estritamente **voluntária**: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação e retirar o consentimento para o tratamento dos seus dados pessoais em qualquer momento, sem ter de prestar qualquer justificação. A retirada de consentimento não afeta a legalidade dos tratamentos anteriormente efetuados com base no consentimento prestado.
Os seus dados pessoais serão anónimos, pois apenas será pedido o cargo e os anos de exercício do mesmo. Os dados serão conservados por 6 meses, após o qual serão destruídos ou anonimizados, garantindo-se o seu anonimato nos resultados do estudo, apenas divulgados para efeitos estatísticos, de ensino, comunicação em encontros ou publicações científicas.
Não existem riscos significativos expectáveis associados à participação no estudo.
O Iscte não divulga ou partilha com terceiros a informação relativa aos seus dados pessoais. O Iscte tem um Encarregado de Proteção de Dados, contactável através do email *dpo@iscte-iul.pt*. Caso considere necessário tem ainda o direito de apresentar reclamação à autoridade de controlo competente – Comissão Nacional de Proteção de Dados. **Declaro** ter compreendido os objetivos de quanto me foi proposto e explicado pelo/a investigador/a, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora.

1. Aceito participar no estudo e consinto que os meus dados pessoais sejam utilizados de acordo com a informações que me foram disponibilizadas. *

Sim

Não

FIGURE A.3. Section 1 of Survey

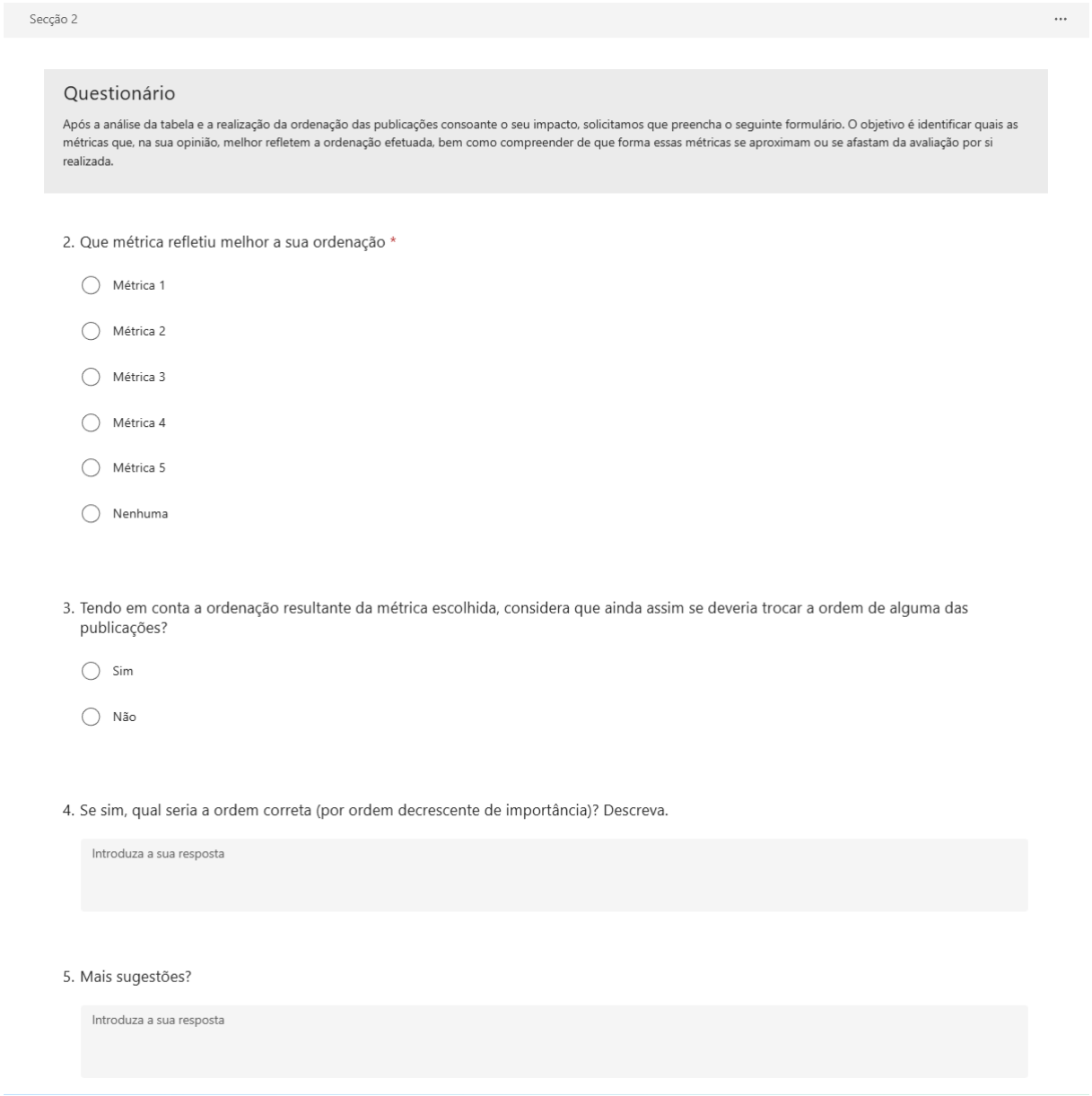


FIGURE A.4. Section 2 of Survey

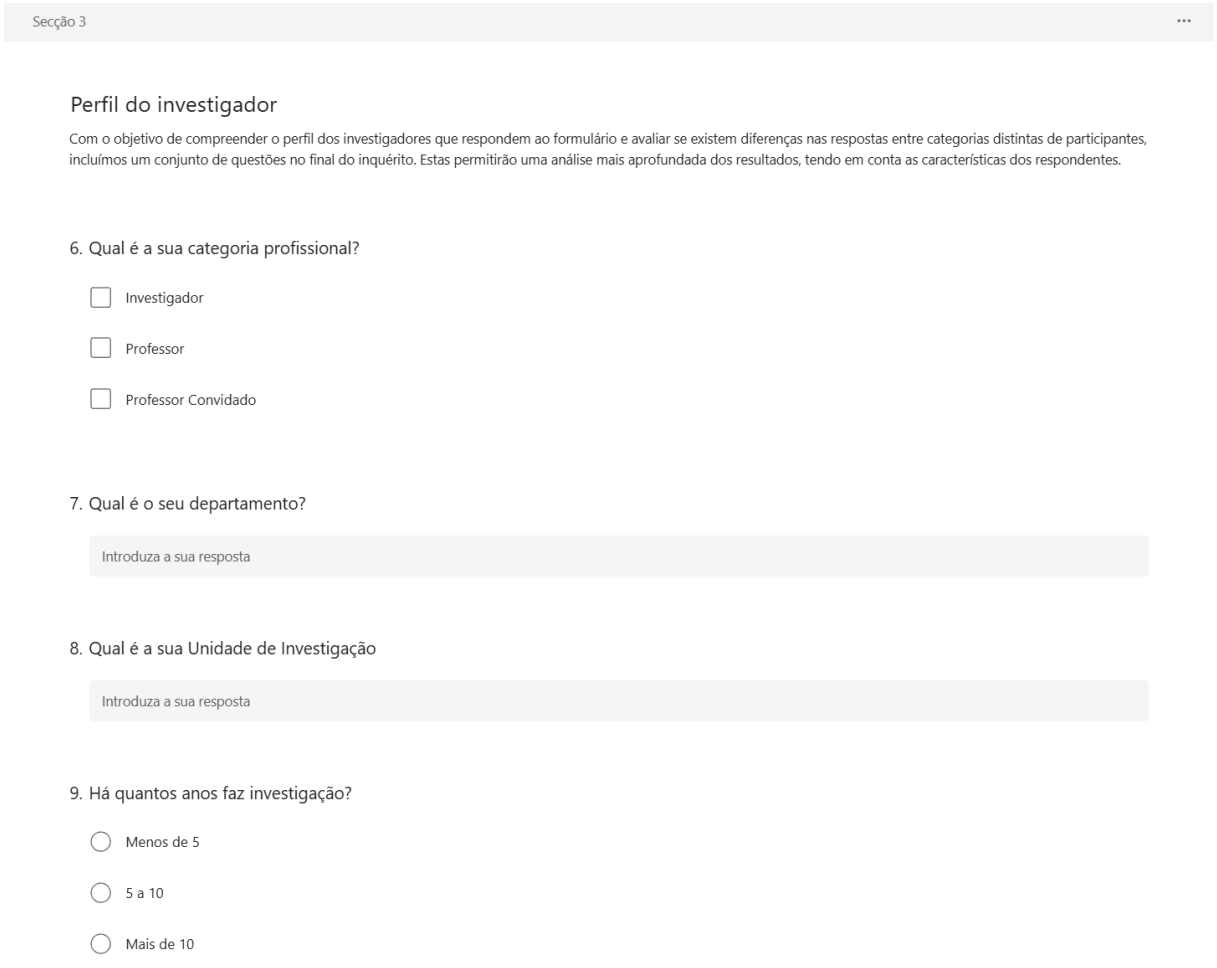


FIGURE A.5. Section 3 of Survey

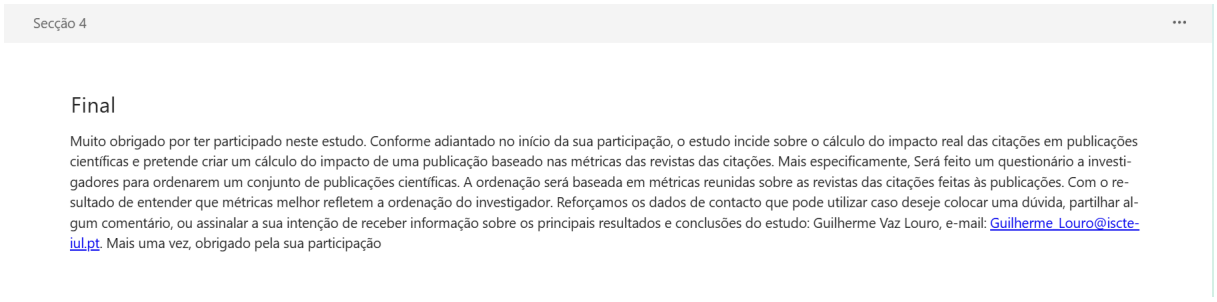


FIGURE A.6. Section 4 of Survey

APPENDIX B

Ethics Committee Acknowledgment

Before initiating any interview procedure, the survey needed to have been approved by the Institution's own Ethics Committee. The approval is shown in the following figures:

COMISSÃO DE ÉTICA DA ESCOLA DE TECNOLOGIAS E ARQUITETURA

PARECER [Final] CE-ISTA/2025.12

Projeto Proj-ISTA/2025.10 "The Calculation of the True Impact of Citations on Scientific Publications

O projeto, submetido por Guilherme Vaz Louro, foi apreciado pela Comissão de Ética da Escola de Tecnologias e Arquitetura.

A informação disponibilizada, no *Formulário de Submissão para Aprovação Ética* em uso no Iscte e anexos associados, satisfaz os requisitos éticos exigíveis neste tipo de projetos de investigação, contemplando, nomeadamente:

a) **O problema de investigação e a relevância do estudo:**

"Falta de um mecanismo de cálculo do impacto de publicações baseado nas citações utilizando métricas publicas não baseadas em contagem de citações. O estudo tem como fim proporcionar um meio para calcular o impacto real de publicações científicas baseado nas métricas de revista disponíveis para as citações.."

b) **O objetivo/perguntas de investigação:**

"Objetivo de criar um cálculo do impacto de uma publicação baseado nas métricas das revistas das citações. É possível calcular o impacto de uma publicação baseado nas métricas das revistas das citações? Qual é a melhor métrica ou métricas para calcular este impacto?."

c) **O método:**

"A equipa de investigação escolherá um universo de participantes que enquadram no âmbito e objetivos do estudo. Será apresentado aos entrevistados um documento de consentimento informado onde a participação ficará confirmada. Será feito um questionário a investigadores para ordenarem um conjunto de publicações científicas. A ordenação será baseada em métricas reunidas sobre as revistas das citações feitas às publicações. Com o resultado de entender que métricas melhor refletem a ordenação do investigador.
Neste questionário será perguntado a categoria profissional: Professor, Investigador ou Professor Convidado, o departamento, a unidade de investigação e há quantos anos exerce o cargo de investigação.
Estes dados serão registados sem ligação a quem os submete, sendo guardados para conhecer a ocupações/função de quem responde e os anos de exercício do cargo, o que garante a anonimidade dos dados. Após 6 meses os dados serão destruídos.."

Iscte – Instituto Universitário de Lisboa • Av. Forças Armadas, 1649-026 Lisboa • ☎ +351 217 903 000 • ✉ geral@iscte-iul.pt

A3ES AACSB ACCREDITED AMBA ACCREDITED EUR-ACE IEP INSTITUTIONAL EVALUATION PROGRAMME Project Management Institute CERTIFIED ISO 9001 ISO 14001 CERTIFIED NP4469

Observações da CE:

- O investigador assinalou haver recolha de dados pessoais, mas afirma que a recolha de dados é anónima, o que não é consistente. Com efeito, afigura-se possível recolher os dados de forma o mais anonimizada possível, apenas recolhendo a categoria profissional, o departamento, a unidade de investigação e os anos de serviço. No entanto, considerando o reduzido tamanho da amostra, e os contactos prévios possivelmente registados (e.g. via email) para o recrutamento, é aconselhável considerar-se haver recolha de dados pessoais, na medida em que o risco de re-identificação não é negligenciável.
- Não se afiguram existir riscos substanciais para os participantes.
- Como há recolha de dados pessoais, devem ser adotadas as medidas técnicas e organizativas obrigatórias em vigor na instituição para tratamento de dados pessoais, nomeadamente as constantes na Secção E1 das orientações aos investigadores sobre proteção de dados pessoais em investigação científica, disponíveis no site do CE. Estas incluem, entre outras, a encriptação e a conservação em servidores do Iscte ou através de serviços de nuvem fornecidos pelo Iscte (e.g. Sharepoint do Office 365). Deve ser evitado, na medida do possível, o uso de computadores não institucionais ou portáteis.
- Todos os membros da equipa sem vínculo contratual com o Iscte, designadamente estudantes, devem assinar o Termo de Responsabilidade e Confidencialidade, devendo este ser junto ao processo submetido à CE. O template está disponível em: https://fenix.iscte-iul.pt/documents/RGPD/Minuta_Termo_Resp_Conf_Aluno_ou_Investig_Sem_Cont_Trab_Iscte.doc.

d) **Participantes e método de recrutamento:** Os investigadores referem que o estudo envolverá cerca de cinco a vinte participantes, convidados pela equipa.

Considerações da CE:

- Considerando que o recrutamento envolve contactos registados com os potenciais participantes, essas mensagens constituem dados pessoais e devem ser eliminadas aquando da destruição ou anonimização dos dados, ou assim que deixarem de ser necessárias.

Iscte – Instituto Universitário de Lisboa • Av. Forças Armadas, 1649-026 Lisboa • ☎ +351 217 903 000 • ✉ geral@iscte-iul.pt
A3ES AACSB ACCREDITED AMBA ACCREDITED EUR-ACE® IEP INSTITUTIONAL EVALUATION PROGRAMME Project Management Institute CERTIFIED ISO 9001 ISO 14001 CERTIFIED NP4469

e) O **consentimento informado, livre e esclarecido** (CILE) adota o modelo em vigor recomendado pela CE para tratamento de dados pessoais. Contudo, carece de correções, designadamente:

- **Deve fazer referência de forma clara às categorias de dados recolhidas.**
- **O seguinte parágrafo deve ser corrigido, pois, assumindo-se a existência de recolha de dados pessoais, não é correto afirmar que "os seus dados pessoais serão anónimos". Na verdade, ainda que se recolham apenas dados indiretos (como cargo, departamento, unidade de investigação e anos de serviço), o reduzido tamanho da amostra e os contactos prévios estabelecidos (por exemplo, por email) para recrutamento tornam plausível a reidentificação:**
  "Os seus dados pessoais serão anónimos, pois apenas será pedido o cargo e os anos de exercício do mesmo. Os dados serão conservados por 6 meses, após o qual serão destruídos ou anonimizados, garantindo-se o seu anonimato nos resultados do estudo, apenas divulgados para efeitos estatísticos, de ensino, comunicação em encontros ou publicações científicas."
  Assim, sugere-se o seguinte ou equivalente:
  "Serão recolhidos apenas dados indiretos, como a categoria profissional, o departamento/unidade de investigação e os anos de serviço. Apesar de não serem recolhidos dados identificativos diretos, a combinação de variáveis e o pequeno tamanho da amostra podem permitir a reidentificação dos participantes, pelo que os dados serão tratados como dados pessoais. Os dados recolhidos serão conservados durante seis meses, após o que serão destruídos ou anonimizados de forma irreversível. Os resultados serão apresentados de forma agregada, garantindo a não identificação dos participantes, e apenas utilizados para fins estatísticos, de ensino, comunicação em encontros ou publicações científicas."
- **Finalmente, recorda-se que os consentimentos são, em si, dados pessoais. Os documentos em papel devem ser digitalizados (por exemplo, um mês após a recolha), eliminando-se os originais em papel. As versões digitais deverão ser conservadas separadamente dos demais dados, até ao momento da anonimização ou destruição dos restantes dados.**

Iscte – Instituto Universitário de Lisboa • Av. Forças Armadas, 1649-026 Lisboa • ☎ +351 217 903 000 • ✉ geral@iscte-iul.pt
A3ES AACSB ACCREDITED AMBA ACCREDITED EUR-ACE® IEP INSTITUTIONAL EVALUATION PROGRAMME Project Management Institute CERTIFIED ISO 9001 ISO 14001 CERTIFIED NP4469

f) O estudo **não** inclui uma amostra proveniente de **populações vulneráveis**.

g) Relativamente ao debriefing, obedece aos requisitos recomendados pela CE.

h) A Declaração de Responsabilidade e de Conduta Ética dos Investigadores está em conformidade com as disposições contidas no Código de Conduta Ética do Iscte.

Em suma, assegurados que se encontram a natureza voluntária da participação, o consentimento livre e informado, medidas adequadas para a proteção de dados pessoais, a confidencialidade dos dados coligidos, entende a Comissão de Ética emitir parecer final favorável à realização da investigação, sem prejuízo das recomendações acima indicadas. Caso os investigadores entendam não implementar as recomendações sugeridas, este parecer é considerado intercalar, devendo os investigadores submeter fundamentação adicional.

Relatores: Nuno David, Catarina Ferreira da Silva

Lisboa, 30 de julho de 2025

O Presidente da Comissão,
Professor Nuno David

A relatora,
Professora Catarina Ferreira da Silva

Assinado por: **Catarina Eufémia Domingues Alves Ferreira da Silva**
Num. de Identificação: 10346329
Data: 2025.07.30 11:13:50 +0100
Certificado por: **Diário da República**
Atributos certificados: **Diretora do Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura (ISTAR) - ISCTE - Instituto Universitário de Lisboa**

Iscte – Instituto Universitário de Lisboa • Av. Forças Armadas, 1649-026 Lisboa • ☎ +351 217 903 000 • ✉ geral@iscte-iul.pt

A3ES AACSB ACCREDITED AMBA ACCREDITED EUR-ACE® IEP INSTITUTIONAL EVALUATION PROGRAMME Project Management Institute CERTIFIED ISO 9001 ISO 14001 CERTIFIED NP4469

FIGURE B.1. Ethics Committee Approval